THE MAXIMUM ENTROPY SPECTRUM AND THE BURG TECHNIQUE

TECHNICAL REPORT NUMBER 1: ADVANCED SIGNAL PROCESSING

TEXAS INSTRUMENTS, INCORPORATED

PREPARED FOR
OFFICE OF NAVAL RESEARCH

25 JUNE 1975

ALEX(83)-TR-75-01

THE MAXIMUM ENTROPY SPECTRUM AND THE BURG TECHNIQUE

TECHNICAL REPORT NO. 1

ADVANCED SIGNAL PROCESSING

Prepared by
Thomas E. Barnard

TEXAS INSTRUMENTS INCORPORATED
Equipment Group
Post Office Box 6015
Dallas, Texas 75222

Sponsored by

NAVAL INTELLIGENCE SUPPORT CENTER
Washington, D.C. 20390

Prepared for

OFFICE OF NAVAL RESEARCH
Arlington, Virginia 22217
Contract Number: N00014-75-C-0101

25 June 1975

THE MAXIMUM ENTROPY SPECTRUM AND THE BURG TECHNIQUE

TECHNICAL REPORT NO. 1

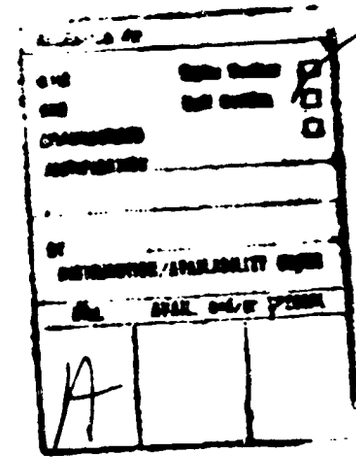
ADVANCED SIGNAL PROCESSING

Prepared by
Thomas E. Barnard

TEXAS INSTRUMENTS INCORPORATED
Equipment Group
Post Office Box 6015
Dallas, Texas 75222

Sponsored by

NAVAL INTELLIGENCE SUPPORT CENTER
Washington, D.C. 20390



Prepared for

OFFICE OF NAVAL RESEARCH
Arlington, Virginia 22217
Contract Number: N00014-75-C-0101

25 June 1975

ERRATA

p. I-1 : Replace

... only when a rectangular taper is applied to the autocorrelation function. Again, in computing conventional spectral estimates, only a rectangular taper preserves the autocorrelation function intact.

with

... only when the available autocorrelation function lag values are Fourier transformed in their original unmodified form. To reduce the spectral window effects associated with this procedure, some conventional spectral analysis methods taper the autocorrelation function.

p. III-2: Replace $\left| p_t^N \left(a_{N+1}^{N+1}\right)^* q_{t+N}^N \right|^2$ with $\left| p_t^N + \left(a_{N+1}^{N+1}\right)^* q_{t+N}^N \right|^2$

p. III-4: Replace $(q_{N+1}, q_{N+2}, \ldots, q_t)$ with $(q_{N+1}, q_{N+2}, \ldots, q_T)$

p. E-5 : Replace $\sum_{m=1}^{M} k^{M-1}$ with $\sum_{m=1}^{M} k^{m-1}$

p. E-5 : Replace $\dfrac{-\Delta t \sum_{m=1}^{M} (m-1)k^{M-1}}{\sum_{m=1}^{M} k^{M-1}}$ with $\dfrac{-\Delta t \sum_{m=1}^{M} (m-1)k^{m-1}}{\sum_{m=1}^{M} k^{m-1}}$

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1 REPORT NUMBER	2 GOVT ACCESSION NO. 3 RECIPIENT'S CATALOG NUMBER
4 TITLE (and Subtitle) THE MAXIMUM ENTROPY SPECTRUM AND THE BURG TECHNIQUE	5 TYPE OF REPORT & PERIOD COVERED Technical 6 PERFORMING ORG. REPORT NUMBER ALEX(03)-TR-75-01
7 AUTHOR(s) Thomas E. Barnard	8 CONTRACT OR GRANT NUMBER(s) N00014-75-C-0101
9 PERFORMING ORGANIZATION NAME AND ADDRESS Texas Instruments Incorporated Equipment Group Dallas, Texas 75222	10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11 CONTROLLING OFFICE NAME AND ADDRESS Naval Intelligence Support Center 4301 Suitland Road Washington, D.C. 20390	12 REPORT DATE 25 June 1975 13 NUMBER OF PAGES 84
14 MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research 800 N. Quincy Street Arlington, Virginia 22217	15 SECURITY CLASS. (of this report) UNCLASSIFIED 15a DECLASSIFICATION/DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum entropy spectrum
Burg technique
Spectral estimation
Time series analysis
Autoregressive spectral estimator
Innovations sequence

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

This tutorial paper describes the maximum entropy spectrum and the Burg technique for computing the prediction error power and prediction error filter coefficients in the associated spectral estimation formula. The maximum entropy spectrum is identical to the autoregressive spectral estimator. Also included in this paper is a discussion of the K-line spectrum, which is the wavenumber analogue of the frequency-domain maximum entropy spectrum, and the Burg technique modifications necessary for its implementation

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Continued

The purpose of this paper is to provide a complete and self-contained account of the main features of the maximum entropy spectrum. Since many of the relevant mathematical derivations are not found in the formal published literature, they are incorporated in this paper. Supporting material and various sidelights of the maximum entropy spectrum appear in the appendices.

ACKNOWLEDGMENT

Mrs. Bonnie C. Taylor and Mrs. Cherylann B. Saunders typed the text and prepared the figures for this paper.

ABSTRACT

This tutorial paper describes the maximum entropy spectrum and the Burg technique for computing the prediction error power and prediction error filter coefficients in the associated spectral estimation formula. The maximum entropy spectrum is identical to the autoregressive spectral estimator. Also included in this paper is a discussion of the K-line spectrum, which is the wavenumber analogue of the frequency-domain maximum entropy spectrum, and the Burg technique modifications necessary for its implementation.

The purpose of this paper is to provide a complete and self-contained account of the main features of the maximum entropy spectrum. Since many of the relevant mathematical derivations are not found in the formal published literature, they are incorporated in this paper. Supporting material and various sidelights of the maximum entropy spectrum appear in the appendices.

TABLE OF CONTENTS

LIST OF FIGURES

SECTION I
INTRODUCTION

The maximum entropy spectrum is an outgrowth of the deconvolution filtering technique long used in oil-exploration data processing. A deconvolution filter is a whitening filter whose purpose is to sharpen the images of a seismic profile. John Parker Burg in the early 1960's noticed that high-resolution power density spectra could be computed using the reciprocal of the squared amplitude response of the deconvolution filter. This form of spectral estimator was known as the Markov spectrum and is identical to the autoregressive spectral estimator independently developed and described in the statistical literature. Later Burg recognized that the Markov spectrum is the maximum entropy spectrum (Burg, 1967) of all possible power spectra agreeing with the measured autocorrelation function values. In addition, Burg developed a method (Burg, 1968) for directly calculating the coefficients of the deconvolution filter (or prediction error filter) used in the spectral estimate. This method produces more accurate spectra and minimizes the problem of end effects.

Most of the problems encountered in conventional spectral estimation are remedied by the maximum entropy spectrum and the Burg technique for calculating the prediction error filter coefficients. For example, the inverse Fourier transform of the maximum entropy spectrum agrees with the measured autocorrelation function values, whereas this is true in conventional spectral estimation only when a rectangular taper is applied to the autocorrelation function. Again, in computing conventional spectral estimates, only a rectangular taper preserves the autocorrelation function intact. The maximum entropy

spectrum, on the other hand, always uses the autocorrelation function in unmodified form. The Burg technique, furthermore, guarantees that the power spectrum is always positive, that the prediction error filter is minimum phase, and that the autocorrelation matrix corresponding to the prediction error filter coefficients is always non-negative definite. With conventional spectral estimates, a zero extension of the autocorrelation function is implicitly assumed. In some cases, this assumption is unreasonable and produces negative power in the spectral estimate. In all cases, the truncation of the autocorrelation function produces lower resolution than the maximum entropy spectrum, which achieves its increased resolution through an optimal extension of the autocorrelation function.

The purpose of this survey paper is to discuss the important features of the maximum entropy spectrum and the Burg technique and to present the relevant derivations, which are often not readily available. The exposition relies heavily on Burg's first two published papers. Section II deals with the maximum entropy spectrum, Section III with the Burg technique, and Section IV with the K-line spectrum, which is the wavenumber analogue of the maximum entropy spectrum. The appendices contain supporting material obtained from a number of sources, which are referenced for the reader's benefit.

SECTION II

THE MAXIMUM ENTROPY SPECTRUM

Given an infinite-length discrete time series (possibly complex) with elements $(\ldots x_{-T}, x_{1-T}, \ldots, x_{-1}, x_0, x_1, \ldots, x_{T-1}, x_T, \ldots)$ sampled at the time interval Δt, the associated NXN autocorrelation matrix is

$$R_N = \overline{\begin{bmatrix} x_t \\ x_{t+1} \\ \vdots \\ x_{t+N-1} \end{bmatrix} \begin{bmatrix} x_t^* & x_{t+1}^* & \cdots & x_{t+N-1}^* \end{bmatrix}}$$

$$= \begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \end{bmatrix},$$

where the vinculum denotes averaging over time, the asterisk denotes complex conjugate, and the elements

$$r(j) = \lim_{T \to \infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x_t x_{t+j}^* \qquad (1-N \le j \le N-1)$$

of the matrix are the autocorrelation function values at the lags $j\Delta t$. The autocorrelation matrix is Hermitian ($R_N = R_N^H$, where the superscript H denotes conjugate transpose), Töplitz (having identical elements along each diagonal), and non-negative definite ($V^H R_N V \ge 0$ for all non-zero N-component column vectors V).

Of all possible power spectra $P(f)$ whose inverse Fourier transforms

$$\int_{-W/2}^{W/2} P(f)\, e^{i2\pi fj\Delta t}\, df$$

agree with the available autocorrelation function values $r(j)$ from $j = 1-N$ to $j = N-1$, the maximum entropy spectrum (Burg, 1967) is the spectrum maximizing the integral

$$\int_{-W/2}^{W/2} \log P(f)\, df \ ,$$

which is proportional to the entropy of a Gaussian band-limited time series with power spectrum $P(f)$ and bandwidth $W = 1/\Delta t$ corresponding to the sampling interval Δt. Entropy maximization subject to the specified constraints on the power spectrum occurs when

$$\int_{-W/2}^{W/2} \log P(f)\, df - \sum_{j=1-N}^{N-1} b_j \left[\int_{-W/2}^{W/2} P(f)\, e^{i2\pi fj\Delta t}\, df - r(j) \right]$$

reaches its maximum value through the proper choice of the power spectrum $P(f)$. The quantities b_j $(j = 1-N, \ldots, -1, 0, 1, \ldots, N-1)$ are complex-valued Lagrangian multipliers. To satisfy the extremal condition, the partial derivative of the integrand with respect to the power spectrum must be zero:

$$0 = \frac{\partial}{\partial P(f)} \left[\log P(f) - \sum_{j=1-N}^{N-1} b_j\, P(f)\, e^{i2\pi fj\Delta t} \right]$$

$$= \frac{1}{P(f)} - \sum_{j=1-N}^{N-1} b_j e^{i2\pi f j \Delta t} .$$

Thus the form of the maximum entropy spectrum is

$$P(f) = \frac{1}{\sum_{j=1-N}^{N-1} b_j e^{i2\pi f j \Delta t}} = \frac{1}{\sum_{j=1-N}^{N-1} b_j z^{-j}} ,$$

where $z = e^{-i2\pi f \Delta t}$. To make the power spectrum real-valued and to satisfy the conditions $r(-j) = r^*(j)$, the Lagrangian multiplier b_{-j} must be the complex conjugate b_j^* of the multiplier b_j whose subscript has the opposite sign.

The equation

$$\sum_{j=1-N}^{N-1} b_j z^{-j} = 0$$

has 2(N-1) roots. If z is a root, then $(z^*)^{-1}$ is also a root: if

$$b_{N-1}^* z^{N-1} + b_{N-2}^* z^{N-2} + \ldots + b_1^* z + b_o + b_1 z^{-1}$$

$$+ \ldots + b_{N-2} z^{-(N-2)} + b_{N-1} z^{-(N-1)} = 0 ,$$

then

$$b_{N-1}\left(\frac{1}{z^*}\right)^{-(N-1)} + b_{N-2}\left(\frac{1}{z^*}\right)^{-(N-2)} + \ldots + b_1\left(\frac{1}{z^*}\right)^{-1} + b_c$$

$$+ b_1^*\left(\frac{1}{z^*}\right)^{1} + \ldots + b_{N-2}^*\left(\frac{1}{z^*}\right)^{N-2}$$

$$+ b_{N-1}^*\left(\frac{1}{z^*}\right)^{N-1} = 0$$

Thus the number of roots for which $|z|>|$ is equal to the number of roots for which $|z|<|$. If no roots lie on the unit circle, there are N-1 roots outside the unit circle and N-1 roots inside the unit circle. The denominator of the maximum entropy spectrum can then be written

$$\frac{W}{P_N}\left[1 + a_2 z^{-1} + a_3 z^{-2} + \ldots + a_N z^{-(N-1)}\right]\left[1 + a_2^* z\right.$$

$$\left. + a_3^* z^2 + \ldots + a_N^* z^{N-1}\right] \quad ,$$

where P_N is real. All of the roots within the unit circle are incorporated in the factor at the left. Thus the maximum entropy spectrum is equal to

$$\frac{P_N}{W}\left(\frac{1}{V^H A A^H V}\right),$$

where the N-component column vectors V and A are, respectively,

$$V = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{N-1} \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-i2\pi f\Delta t} \\ \vdots \\ e^{-i2\pi f(N-1)\Delta t} \end{bmatrix} \text{ and } A = \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} .$$

The constraints embodied in the matrix equation

$$\int_{-W/2}^{W/2} P(f)\, VV^H\, df = \frac{P_N}{W} \int_{-W/2}^{W/2} \frac{VV^H}{V^H AA^H V}\, df = R_N$$

$$= \begin{bmatrix} r(0) & r(1) \cdots\cdots r(N-1) \\ r^*(1) & r(0) \cdots\cdots r(N-2) \\ \vdots & \vdots \qquad \ddots \qquad \vdots \\ r^*(N-1) & r^*(N-2) \cdots\cdots r(0) \end{bmatrix}$$

determine P_N and the components $a_2, a_3, \ldots, a_N$ of the column vector A. Multiplication on the right by the column vector A yields

$$\frac{P_N}{W} \int_{-W/2}^{W/2} \frac{V(V^H A)}{(V^H A)(A^H V)}\, df = R_N A$$

or

$$\frac{P_N}{i2\pi} \oint \frac{V}{zA^H V}\, dz = R_N A \quad \left(dz = \frac{-i2\pi z}{W}\, df \right) ,$$

where the contour-integral path runs counterclockwise along the unit circle. (Reversal of the clockwise path $z = e^{-i2\pi f\Delta t}$ corresponding to increasing frequency eliminates the minus sign in the expression for dz.) The j-th component of $R_N A$ is

$$\frac{P_N}{i2\pi} \oint \frac{z^{j-2} dz}{A^H V} \quad .$$

Since the factor $V^H A$ contains all of the roots within the unit circle, the other factor $A^H V$ is never equal to zero within the unit circle. For all components j from 2 to N, the power of z is non-negative and the integrand contains no poles within the unit circle. For $j = 1$, however, there is a simple pole at $z = 0$ whose residue is

$$\lim_{z \to 0} z \left[\frac{1}{z\left(1 + a_2^* z + a_3^* z^2 + \ldots + a_N^* z^{N-1}\right)} \right] = 1.$$

Canchy's residue theorem implies, therefore, that the first component of $R_N A$ is P_N and that the remaining components from the second through the N-th are zero. Thus the maximum entropy spectrum is

$$P(f) = \frac{P_N / W}{V^H A A^H V} = \frac{P_N \Delta t}{\left| 1 + \sum_{j=1}^{N-1} a_{j+1} e^{i2\pi f j \Delta t} \right|^2} \quad ,$$

where P_N and $a_2, a_3, \ldots, a_N$ are solutions of the matrix equation

$$\begin{bmatrix} r(0) & r(1) \cdots\cdots r(N-1) \\ r^*(1) & r(0) \cdots\cdots r(N-2) \\ \vdots & \vdots \\ r^*(N-1) & r^*(N-2) \cdots r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} P_N \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

Rearrangement of the bottom N-1 rows produces

$$\begin{bmatrix} r(0) & r(1) \cdots\cdots r(N-2) \\ r^*(1) & r(0) \cdots\cdots r(N-3) \\ \vdots & \vdots \\ r^*(N-2) & r^*(N-3) \cdots r(0) \end{bmatrix} \begin{bmatrix} -a_2 \\ -a_3 \\ \vdots \\ -a_N \end{bmatrix} = \begin{bmatrix} r^*(1) \\ r^*(2) \\ \vdots \\ r^*(N-1) \end{bmatrix}$$

or

$$\begin{bmatrix} x^*_{t-1} \\ x^*_{t-2} \\ \vdots \\ x^*_{t-(N-1)} \end{bmatrix} \begin{bmatrix} x_{t-1} & x_{t-2} \cdots\cdots x_{t-(N-1)} \end{bmatrix} \begin{bmatrix} -a_2 \\ -a_3 \\ \vdots \\ -a_N \end{bmatrix} = \begin{bmatrix} x^*_{t-1} \\ x^*_{t-2} \\ \vdots \\ x^*_{t-(N-1)} \end{bmatrix} x_t ,$$

the design equation for an optimum (N-1)-point-long forward prediction filter with weights ($-a_2$, $-a_3$, . . . , $-a_N$), as shown in Figure II-1(a). Taking the complex conjugate yields

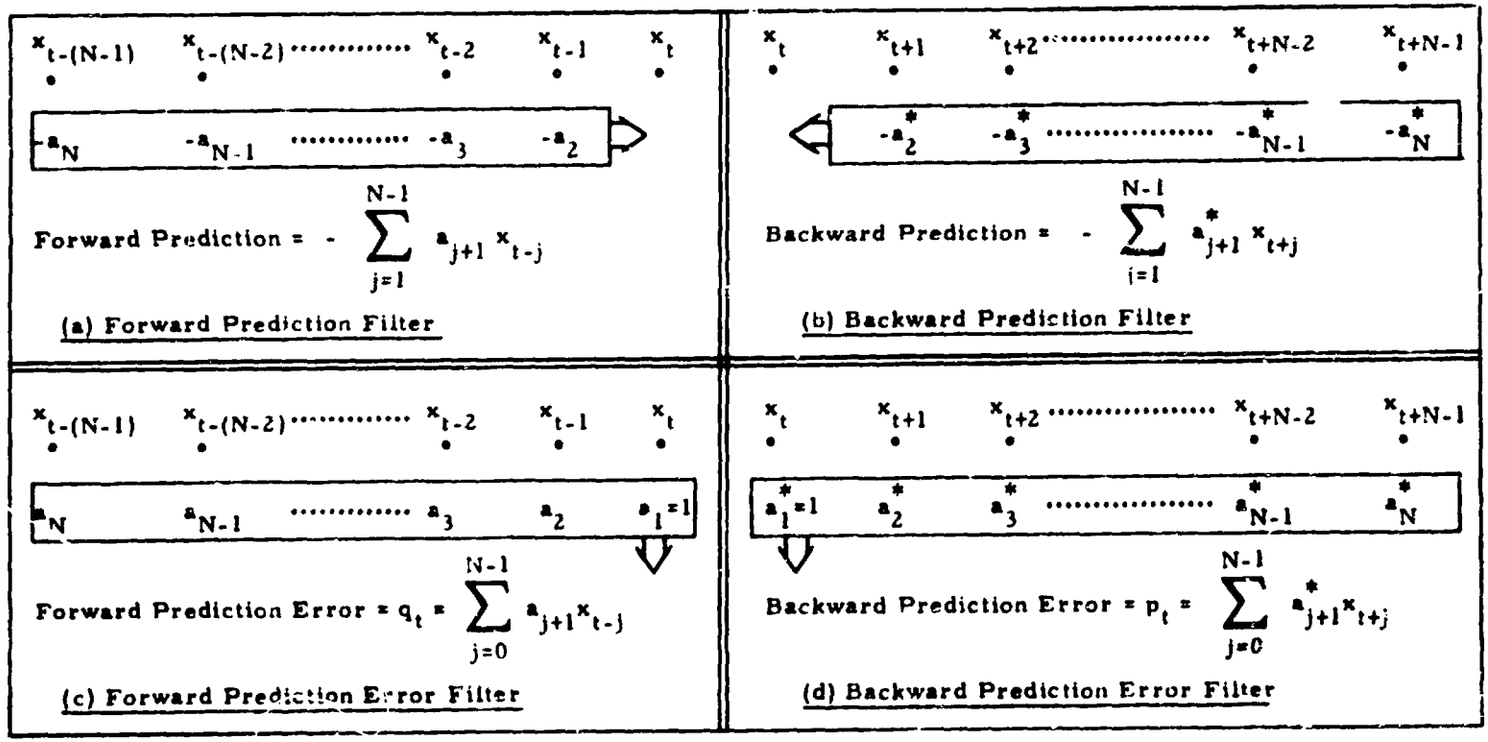


FIGURE II-1

RELATIONSHIP BETWEEN PREDICTION FILTER
AND PREDICTION ERROR FILTER

$$\begin{bmatrix} r(0) & r^*(1) & \cdots & r^*(N-2) \\ r(1) & r(0) & \cdots & r^*(N-3) \\ \vdots & \vdots & & \vdots \\ r(N-2) & r(N-3) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} -a_2^* \\ -a_3^* \\ \vdots \\ -a_N \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(N-1) \end{bmatrix}$$

or

$$\overline{\begin{bmatrix} x_{t+1}^* \\ x_{t+2}^* \\ \vdots \\ x_{t+N-1}^* \end{bmatrix} \begin{bmatrix} x_{t+1} & x_{t+2} & \cdots & x_{t+N-1} \end{bmatrix}} \begin{bmatrix} -a_2^* \\ -a_3^* \\ \vdots \\ -a_N^* \end{bmatrix} = \overline{\begin{bmatrix} x_{t+1}^* \\ x_{t+2}^* \\ \vdots \\ x_{t+N-1}^* \end{bmatrix}} x_t ,$$

the design equation for an optimum (N-1)-point-long backward prediction filter with weights $(-a_N^*, \ldots, -a_3^*, -a_2^*)$, as shown in Figure II-1(b). If the prediction filter output is subtracted from the actual value x_t, the result is the prediction error, which can be produced from a filter with unity weight at zero lag and the negative of the prediction filter weights at their respective lags. Figures II-1(c) and II-1(d) illustrate the forward and backward prediction error filters. The term P_N in the maximum entropy spectrum formula is the mean square error of the prediction filter and the power of the prediction error filter output:

$$P_N = \begin{bmatrix} 1 & a_2^* & \cdots & a_N^* \end{bmatrix} \begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

$$= \begin{bmatrix} 1 & a_2^* \cdots\cdots a_N^* \end{bmatrix} \overline{\begin{bmatrix} x_t \\ x_{t+1} \\ \vdots \\ x_{t+N-1} \end{bmatrix} \begin{bmatrix} x_t^* & x_{t+1}^* \cdots\cdots x_{t+N-1}^* \end{bmatrix}} \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

$$= \overline{\left| x_t - \sum_{j=1}^{N-1} (-a_{j+1}^*)\, x_{t+j} \right|^2} = \overline{p_t p_t^*}$$

$$= \begin{bmatrix} 1 & a_2^* \cdots\cdots a_N^* \end{bmatrix} \overline{\begin{bmatrix} x_t^* \\ x_{t-1}^* \\ \vdots \\ x_{t-(N-1)}^* \end{bmatrix} \begin{bmatrix} x_t & x_{t+1} \cdots\cdots x_{t-(N-1)} \end{bmatrix}} \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

$$= \overline{\left| x_t - \sum_{j=1}^{N-1} (-a_{j+1})\, x_{t-j} \right|^2} = \overline{q_t q_t^*}$$

Inverse Fourier transformation of the maximum entropy spectrum produces autocorrelation function estimates at lags greater than N-1. Addition of the element $z^N = e^{-i2\pi fN\Delta t}$ to the vector V permits the (N+1) X (N+1) autocorrelation matrix to be determined:

$$\frac{P_N}{W}\int_{-W/2}^{W/2} \frac{\begin{bmatrix} V \\ z^N \end{bmatrix}\begin{bmatrix} V & z^{-N} \end{bmatrix}}{\begin{bmatrix} V^H & z^{-N} \end{bmatrix}\begin{bmatrix} A \\ 0 \end{bmatrix}\begin{bmatrix} A^H & 0 \end{bmatrix}\begin{bmatrix} V \\ z^N \end{bmatrix}} df =$$

$$R_{N+1} = \left[\begin{array}{ccc|c} r(0) & r(1) \cdots\cdots\cdots & r(N-1) & r(N) \\ r^*(1) & r(0) \cdots\cdots\cdots & r(N-2) & r(N-1) \\ \vdots & \ddots & \vdots & \vdots \\ r^*(N-1) & r^*(N-2) \cdots\cdots & r(0) & r(1) \\ \hline r^*(N) & r^*(N-1) \cdots\cdots & r^*(1) & r(0) \end{array}\right]$$

Multiplication on the right by the extended prediction error filter column vector and transformation to the z-plane yields

$$\frac{P_N}{i2\pi}\oint \frac{\begin{bmatrix} V \\ N \end{bmatrix}}{zA^H V} dz = R_{N+1}\begin{bmatrix} A \\ 0 \end{bmatrix} ,$$

where the contour path runs counterclockwise along the unit circle. The (N+1)-st component is zero since all poles of the integrand $z^{N-1}/(A^H V)$ lie outside the unit circle. The value $r^*(N)$ can be determined from the equation for the (N+1)-st component. Repeated application of this procedure yields the values $r^*(N+1)$, $r^*(N+2)$, etc., from the matrix equation

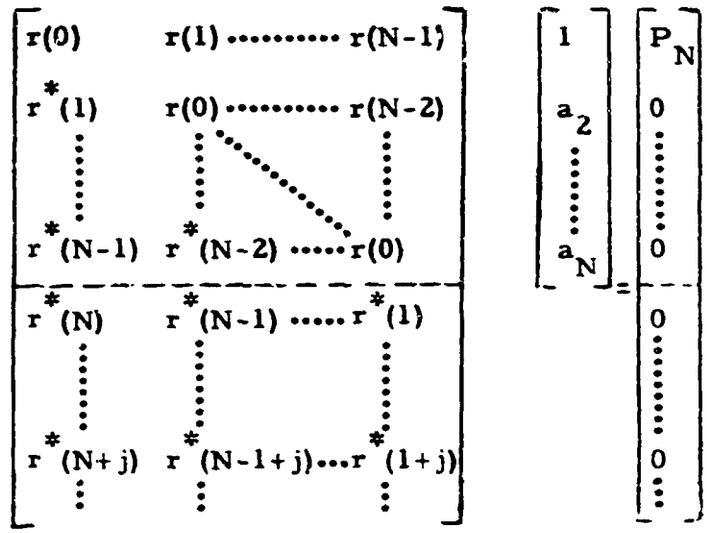


Figure II-2(b) illustrates how the backward extension of the autocorrelation function is accomplished using the complex conjugate of the backward prediction filter. Figure II-2(a) depicts the corresponding forward extension with the complex conjugate of the forward prediction filter.

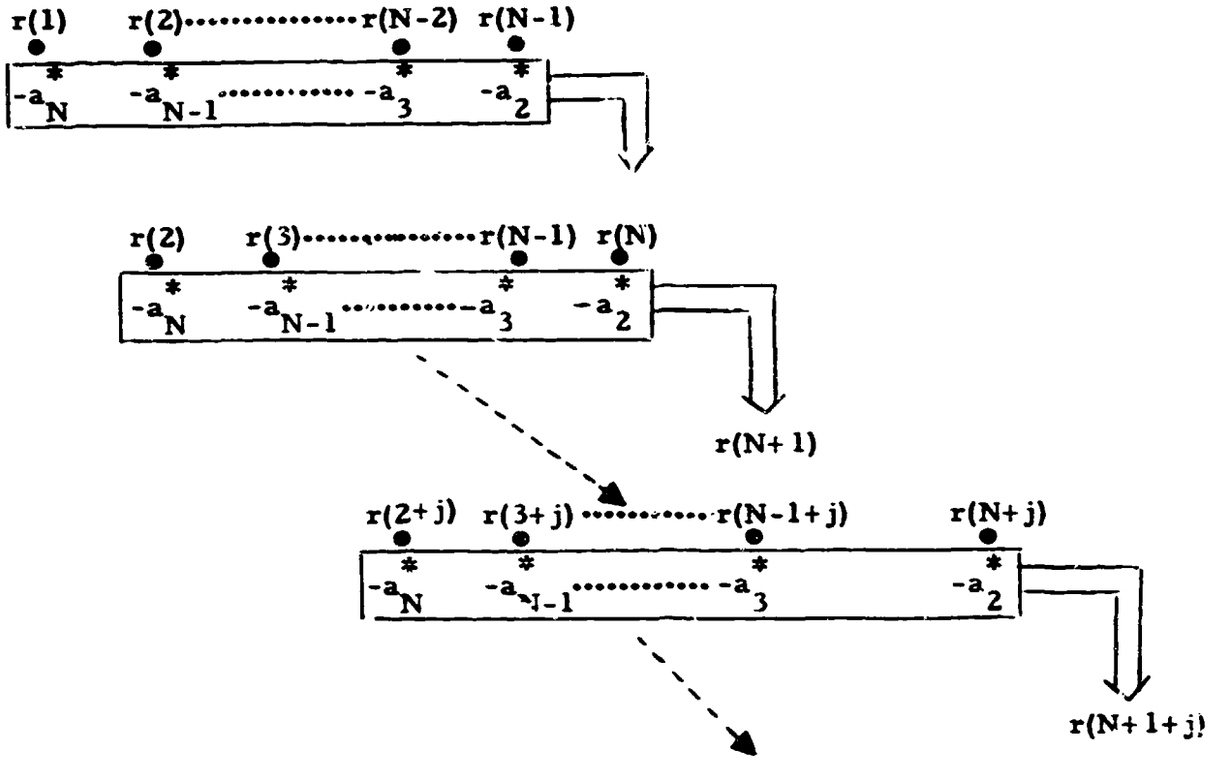
When this maximum entropy method of autocorrelation function extension determines the unavailable lag values, the resultant power spectrum

$$P(f) = \frac{1}{W} \sum_{j=-\infty}^{\infty} r(j)\, z^{j} = \frac{1}{W} \sum_{j=-\infty}^{\infty} r(j)\, e^{-i2\pi f j \Delta t}$$

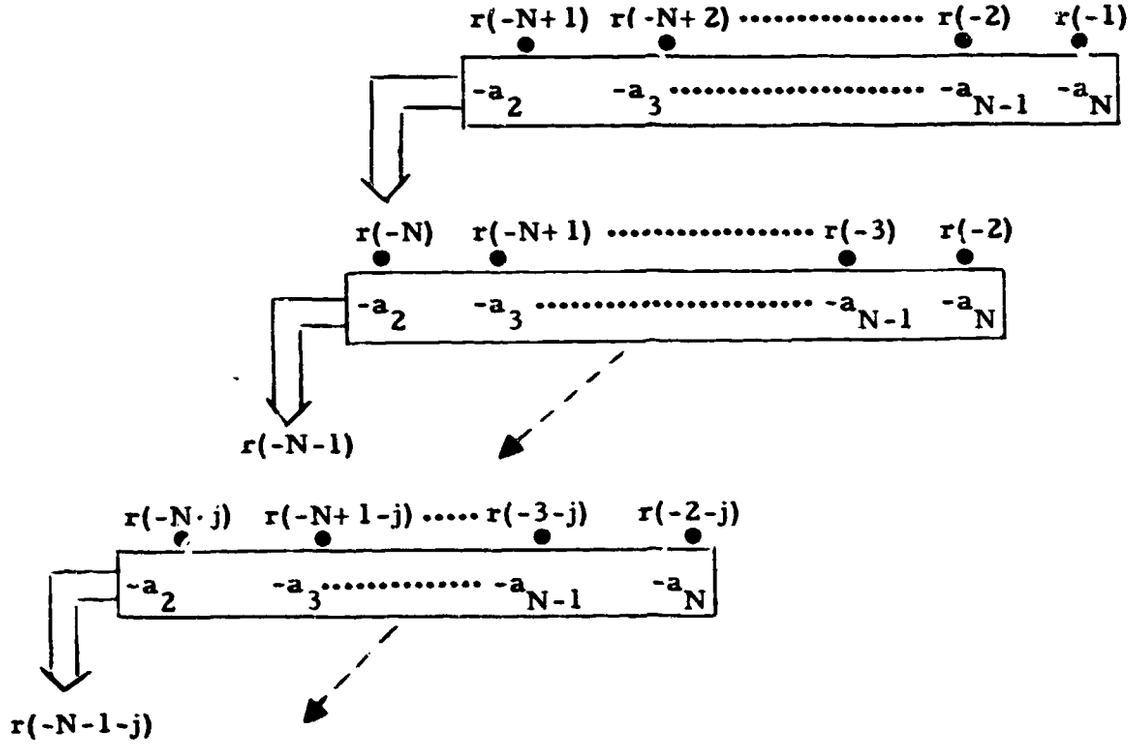
is the maximum entropy spectrum. This fact can be verified by premultiplying the power spectrum by $W\,(V^{H}A)$ and applying the method of undetermined coefficients to the coefficients of $z^{0}, z^{-1}, \ldots, z^{-(N-1)}, z^{-N}, \ldots, z^{-(N+j)}, \ldots$:

$$W(V^{H}A)\, P(f) = V^{H}A \sum_{j=-\infty}^{\infty} r(j)\, z^{j} = \frac{P_N}{A^{H}V} \quad .$$

Conventional spectral estimates with taper functions (Bartlett, Hanning, Hamming, Parzen, etc.) are exactly equal to the convolution of the maximum entropy spectrum with the frequency window corresponding to the particular tapering method employed:



(a) Forward Extension



(b) Backward Extension

FIGURE II-2

EXTENSION OF THE AUTOCORRELATION FUNCTION

$$\int_{f-W/2}^{f+W/2} \left[\frac{1}{W} \sum_{j=(N-1)}^{N-1} w_j e^{-i2\pi\phi j\Delta t} \right] \left[\frac{1}{W} \sum_{k=-\infty}^{\infty} r(k) e^{-i2\pi(f-\phi)k\Delta t} \right] d\phi$$

$$= \frac{1}{W^2} \sum_{j=-(N-1)}^{N-1} \sum_{k=-\infty}^{\infty} w_j r(k) \int_{f-W/2}^{f+W/2} e^{-i2\pi fk\Delta t} \; e^{-i2\pi\phi(j-k)\Delta t} \; d\phi$$

$$= \frac{1}{W^2} \sum_{j=-(N-1)}^{N-1} \sum_{k=-\infty}^{\infty} w_j r(k) e^{-i2\pi fk\Delta t} \; W\delta_{jk}$$

$$= \frac{1}{W} \sum_{j=-(N-1)}^{N-1} w_j r(j) e^{-i2\pi fj\Delta t} \quad ,$$

since

$$\int_{\phi_o}^{\phi_o + W} e^{-i2\pi\phi(j-k)\Delta t} d\phi = W\delta_{jk},$$

where j and k are integers, δ_{jk} is the Kronecker delta operator, and w_j is the taper weighting for the autocorrelation function r(j).

The equations

$$0 = \sum_{j=0}^{N-1} a_{j+1} \, r(j-k) \qquad \text{(for all positive integers k)}$$

in the autocorrelation function extension matrix equation imply that

$$0 = \sum_{j=0}^{N-1} a_{j+1} \overline{x_{t-j} x_{t-k}^*} = \overline{\left[\sum_{j=0}^{N-1} a_{j+1} x_{t-j}\right] x_{t-k}^*} = \overline{q_t x_{t-k}^*}$$

or that the crosscorrelation function between the forward prediction error filter output q_t and the time series values x_{t-k} before the filter output time is zero for all positive values of k. Similarly, the equations

$$0 = \sum_{j=0}^{N-1} a_{j+1} \overline{x_{t+j}^* x_{t+k}} = \overline{\left[\sum_{j=0}^{N-1} a_{j+1} x_{t+j}^*\right] x_{t+k}} = \overline{x_{t+k} p_t^*}$$

imply that the crosscorrelation function between the backward prediction error filter output p_t and the time series values x_{t+k} after the filter output time is also zero for all positive values of k. For the values of k from 1 to N-1, this result is a consequence of the fact that the (N-1)-point-long prediction filter is optimum. For the values of k greater than or equal to N, this result is a consequence of the maximum entropy assumption. If this assumption is correct, the (N-1)-point-long prediction filter is also the infinitely long optimum prediction filter, and the predictability of x_t cannot be improved by using the time series values $x_{t\pm k}$ at lags k whose absolute value is N or greater. If the predictability of x_t can be improved, the entropy or uncertainty of the time series x_t is less than

$$\int_{-W/2}^{W/2} \log \frac{P_N \Delta t}{V^H A A^H V} \, df.$$

Since the forward prediction error filter output q_t is a linear combination of earlier time series values x_{t-k} and the backward prediction error filter output p_t is a linear combination of later time series values x_{t+k}, the equations

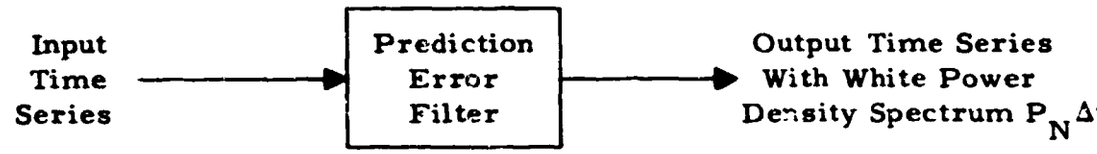
$$\overline{q_t q_{t-j}^*} = \overline{q_t \left[\sum_{k=0}^{N-1} a_{k+1}^* \, x_{t-j-k}^* \right]} = \sum_{k=0}^{N-1} a_{k+1}^* \, \overline{q_t \, x_{t-j-k}^*} = 0 \quad (j \geq 1)$$

$$\overline{p_{t+j} \, p_t^*} = \overline{\left[\sum_{k=0}^{N-1} a_{k+1}^* \, x_{t+j+k} \right] p_t^*} = \sum_{k=0}^{N-1} a_{k+1}^* \, \overline{x_{t+j+k} \, p_t^*} = 0 \quad (j \geq 1)$$

$$\overline{q_t q_t^*} = \overline{q_t \left[\sum_{k=0}^{N-1} a_{k+1}^* \, x_{t-k}^* \right]} = \sum_{k=0}^{N-1} a_{k+1}^* \, \overline{q_t \, x_{t-k}^*} = \overline{q_t \, x_t^*} = P_N$$

$$\overline{p_t p_t^*} = \overline{\left[\sum_{k=0}^{N-1} a_{k+1}^* \, x_{t+k} \right] p_t^*} = \sum_{k=0}^{N-1} a_{k+1}^* \, \overline{x_{t+k} \, p_t^*} = \overline{x_t \, p_t^*} = P_N$$

imply that the prediction error filter output has a white power spectrum, that the filter output power is P_N, and that the filter output power density is P_N/W or $P_N \Delta t$. Figure II-3 portrays the relationship between the input power spectrum and the white prediction error filter output power spectrum corresponding to the maximum entropy assumption. The maximum entropy assumption that the prediction error filter output is white is equivalent to an autoregression model of order N-1 for a second-order zero-mean stationary time series. The maximum entropy spectrum is known as an autoregressive spectrum in the statistical literature (e.g. : Jones, 1974; Gersh and Sharpe, 1973). The assumed whiteness of the forward and backward prediction error filter outputs provides a means of testing the maximum entropy assumption: if the spectra of the prediction error filter outputs are not approximately white, the maximum entropy spectrum is not a good spectral estimate. This fact is useful in determining the best length for the prediction error filter. More formal



$$\text{Input Power Spectrum} = \frac{\text{Output Power Spectrum}}{\text{Power Response of Prediction Error Filter}}$$

$$P(f) = \frac{P_N \Delta t}{\left| 1 + \sum_{j=1}^{N-1} a_{j+1} \, e^{i 2 \pi f j \Delta t} \right|^2}$$

FIGURE II-3

ASSUMED RELATIONSHIP BETWEEN PREDICTION ERROR FILTER OUTPUT POWER SPECTRUM AND INPUT POWER SPECTRUM

methods of identifying the order of the autoregression model such as Akaike's final prediction error (FPE) criterion (Akaike, 1969; Akaike, 1970a; Akaike, 1970b; Akaike, 1971a) and Akaike's information criterion (AIC) (Akaike, 1971b; Akaike, 1972a; Akaike, 1972b; Akaike, 1974) provide a means of objectifying the maximum entropy spectrum. These methods are apparently not widely known in the geophysical literature. Likewise, methods of estimating the spectral mean and variance are available in the statistical literature (Kromer, 1969).

Since all zeroes of

$$A^H V = 1 + \sum_{j=1}^{N-1} a_{j+1}^* z^j$$

lie outside the unit circle, all zeroes of the forward prediction error filter z-transform

$$A(z) = 1 + \sum_{j=1}^{N-1} a_{j+1} z^j$$

also lie outside the unit circle because the roots of A(z) are simply the complex conjugates of the roots of $A^H V$. Thus the forward prediction error filter is a minimum phase filter. Since its output does not precede any of its input points, it is also a causal filter. The z-transform

$$C(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{j=1}^{N-1} a_{j+1} z^j} = 1 + \sum_{j=1}^{\infty} c_{j+1} z^j$$

of the inverse to the forward prediction error filter can be expanded in a power series with no negative powers of z since the forward prediction error filter has no zeroes on or inside the unit circle. The inverse of the forward

prediction error filter is also a causal minimum phase filter. The inverse filter can be used to construct the time series x_t:

$$\sum_{k=0}^{\infty} c_{k+1}\, q_{t-k} = \sum_{k=0}^{\infty} c_{k+1} \left[\sum_{j=0}^{N-1} a_{j+1}\, x_{t-j-k} \right]$$

$$= \sum_{m=0}^{\infty} \left[\sum_{n=0}^{\min(m, N-1)} c_{m-n-1}\, a_{n+1} \right] x_{t-m}$$

$$= \sum_{m=0}^{\infty} \delta_{om}\, x_{t-m} = x_t$$

since

$$1 = \left(\sum_{j=0}^{N-1} a_{j+1}\, z^j \right) \left(\sum_{k=0}^{\infty} c_{k+1}\, z^k \right) = \sum_{m=0}^{\infty} \left[\sum_{n=0}^{\min(m, N-1)} c_{m-n-1}\, a_{n+1} \right] z^m$$

$$= \sum_{m=0}^{\infty} \delta_{om}\, z^m \quad .$$

Thus the input time series can be generated from the forward prediction error filter output using a causal filter (the inverse of the forward prediction error filter). The process is causally invertible: the causal forward prediction error filter produces the forward prediction error filter output. If the maximum

entropy assumption is satisfied (i.e., if the time series q_t is white), then the forward prediction error filter output is the innovations sequence (Kailath, 1970; Kailath, 1968; Kailath and Frost, 1968; Frost and Kailath, 1971; Kailath and Geesey, 1971; Kailath and Geesey, 1973; Gevers and Kailath, 1973; Aasnaes and Kailath, 1973) corresponding to the time series x_t when it is a second-order zero-mean stationary time series. Figure II-4 diagrams the relationship between the innovations sequence and the input time series. An equally apt and more succinct name for a prediction error filter is an innovations filter (as shown in Figure II-4). The backward prediction error filter is a maximum phase filter (a minimum phase filter if the direction of time is reversed). When the maximum entropy assumption is valid, it generates a backward innovations sequence from present and future values of the time series x_t. In the vast majority of practical geophysical processing situations, the optimum prediction error filter coefficients tend to zero as the length of the filter increases. Consequently, the prediction error filter output approaches an innovations sequence as the filter length increases indefinitely.

To solve the prediction error filter design matrix equation, a simplified form of a recursive algorithm first developed by Norman Levinson (Levinson, 1947) and later presented in the statistical literature (Durbin, 1960) is available. The (N+1)-point-long prediction error filter can be created from the N-point-long prediction error filter through the determination of the single coefficient a_{N+1}^{N+1} in the (N+1)-point-long filter:

$$\begin{bmatrix} r(0) & r(1)\cdots\cdots r(N-1) & r(N) \\ r^*(1) & r(0)\cdots\cdots r(N-2) & r(N-1) \\ \vdots & \vdots & \vdots \\ r^*(N-1) & r^*(N-2)\cdots\cdots r(0) & r(1) \\ r^*(N) & r^*(N-1)\cdots\cdots r^*(1) & r(0) \end{bmatrix} \left(\begin{bmatrix} 1 \\ a_2^N \\ \vdots \\ a_N^N \\ 0 \end{bmatrix} + a_{N+1}^{N+1} \begin{bmatrix} 0 \\ (a_N^N)^* \\ \vdots \\ (a_2^N)^* \\ 1 \end{bmatrix} \right)$$

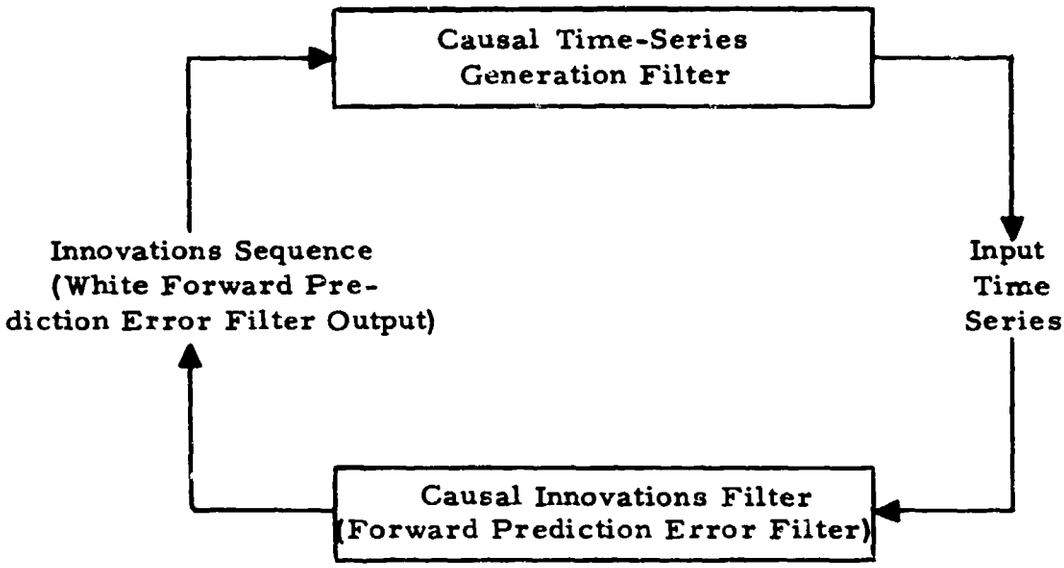


FIGURE II-4

RELATIONSHIP BETWEEN THE INNOVATIONS SEQUENCE
AND THE INPUT TIME SERIES

$$= \left\{ \begin{bmatrix} P_N \\ 0 \\ \vdots \\ 0 \\ \Delta_N \end{bmatrix} + a_{N+1}^{N+1} \begin{bmatrix} \Delta_N^* \\ 0 \\ \vdots \\ 0 \\ P_N \end{bmatrix} \right\} = \begin{bmatrix} P_{N+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

where

$$\Delta_N = \sum_{j=0}^{N-1} a_{j+1}^{N} \; r^*(N-j) \; .$$

Setting

$$a_{N+1}^{N+1} = -\Delta_N / P_N$$

yields the prediction filter weights ($\Delta + a_{N+1}^{N+1} P_N = 0$) and the power P_{N+1} of the (N+1)-point-long prediction error filter output:

$$P_N + a_{N+1}^{N+1} \left[- P_N \left(a_{N+1}^{N+1} \right)^* \right] = P_N \left[1 - a_{N+1}^{N+1} \left(a_{N+1}^{N+1} \right)^* \right] = P_{N+1} \; .$$

The NXN solution proceeds recursively from the one-by-one problem, where

$$\left[r(0) \right] \left[a_1^1 = 1 \right] = P_1 \; .$$

The amplitude of the coefficient a_{N+1}^{N+1} must not exceed one if the (N+1) X (N+1) autocorrelation matrix R_{N+1} is non-negative definite. This condition is equivalent to specifying that the prediction error filter's z-transform has no zeroes inside the unit circle (see Appendix A). A survey paper (Kailath, 1974) on linear filtering theory discusses some of the implications of Levinson's algorithm.

Since

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_2^N & 1 & \ddots & \vdots \\ a_3^N & a_2^{N-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ a_N^N & a_{N-1}^{N-1} & \cdots & a_2^2 \end{bmatrix} = \begin{bmatrix} P_N & ? & \cdots & ? \\ 0 & P_{N-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & ? \\ 0 & \cdots & 0 & P_1 \end{bmatrix},$$

the determinant of R_N is

$$\left|R_N\right| = \prod_{j=1}^{N} P_j \quad ,$$

and the inverse of R_N is

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_2^N & 1 & \ddots & \vdots \\ a_3^N & a_2^{N-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ a_N^N & a_{N-1}^{N-1} & \cdots & a_2^2 \end{bmatrix} \begin{bmatrix} P_N^{-1} & 0 & \cdots & 0 \\ 0 & P_{N-1}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_1^{-1} \end{bmatrix} \begin{bmatrix} 1 & \left(a_2^N\right)^* & \left(a_3^N\right)^* & \cdots & \left(a_N^N\right)^* \\ 0 & 1 & \left(a_2^{N-1}\right)^* & \cdots & \left(a_{N-1}^{N-1}\right)^* \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & \left(a_2^2\right)^* \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

because

$$\begin{bmatrix} 1 & \left(a_2^N\right)^* & \left(a_3^N\right)^* & \cdots & \left(a_N^N\right)^* \\ 0 & 1 & \left(a_2^{N-1}\right)^* & \cdots & \left(a_{N-1}^{N-1}\right)^* \\ \vdots & & \ddots & \ddots & \vdots \\ & & & 1 & \left(a_2^2\right)^* \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \ddots & \ddots & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_2^N & 1 & \ddots & \vdots \\ a_3^N & a_2^{N-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ a_N^N & a_{N-1}^{N-1} & \cdots & a_2^2 \quad 1 \end{bmatrix}$$

$$= \begin{bmatrix} P_N & 0 & \cdots & 0 \\ 0 & P_{N-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_1 \end{bmatrix} .$$

(The zeroes to the right of the main diagonal in the last matrix occur because the matrix is Hermitian, and this fact is true since the matrix to the left of the Hermitian autocorrelation matrix is the conjugate transpose of the matrix to its right.)

If w is substituted for r(N) in the (N+1) X(N+1) autocorrelation matrix R_{N+1}, the determinant is

$$\left|R_{N+1}\right| = \left|R_{N-1}\right| \left(c + w_o^* w + w_o w^* - w w^*\right),$$

where c is a term not involving w or w^* and $\left|R_{N-1}\right| w_o$ is the coefficient for w^*. Since the determinant $\left|R_{N+1}\right|$ is also the product of the prediction error powers P_1 through P_{N+1}, it is also equal to

$$\left|R_{N+1}\right| = \left|R_{N-1}\right| \left(r^2 - \left|w - w_o\right|^2\right) = \left|R_{N-1}\right| P_N^2 \left[1 - a_{N+1}^{N+1} \left(a_{N+1}^{N+1}\right)^*\right] .$$

The determinant is maximized when $a_{N+1}^{N+1} = 0$ or, equivalently, when $w = w_o$, so that $r = P_N$ and

$$w_o = - \sum_{j=1}^{N-1} \left(a_{j+1}^{N} \right)^* r(N-j) \quad .$$

As shown earlier, $r(N) = w_o$ is the maximum entropy extension of the autocorrelation function. An alternate but equivalent entropy definition requires that the determinant $\left| R_{N+1} \right|$ be maximized by the proper choice of r(N) in order to maximize the entropy of the time series x_t (McDonough, 1974). The fact that the maximum entropy autocorrelation function extension described here does indeed maximize the determinant $\left| R_{N+1} \right|$ means that the two entropy definitions are consistent. Since the autocorrelation matrix is non-negative definite, the autocorrelation function value r(N) must lie on or within the w-plane circle described parametically by $w_o + e^{i\theta} P_N$. The maximum entropy autocorrelation function extension places the value r(N) precisely at the center of the circle. Any other choice of r(N) is biased in the sense that it adds information not based on the autocorrelation matrix R_N. Thus, when the autocorrelation matrix R_N is precisely known and nothing is known about the autocorrelation function at lags of N or greater, the maximum entropy spectrum is the most reasonable power spectrum possible.

SECTION III
THE PURG TECHNIQUE

The Burg technique (Burg, 1968) is a method for computing the prediction error filter coefficients directly from a finite-length time series x_t ($t = 1, 2, \ldots, T-1, T$). By using the Levinson recursion relationship described earlier, this procedure guarantees that the forward prediction error filter is minimum phase and that the associated autocorrelation matrix is non-negative definite. Minimizing the average power of both the forward and backward prediction error filter outputs to obtain the filter coefficients provides greater reliability in their estimation. In this algorithm, no assumptions about the time series before or after the measured values are made, so that end effects are avoided. The advantages of the Burg technique are particularly pronounced when the number of points in the time series x_t is small.

Since the coefficient of a one-point-long prediction error filter is one, both the forward and backward prediction error filter outputs at time t are equal to x_t. Computation of the prediction error power for the one-point-long filter initializes the Burg technique:

$$P_1 = \frac{1}{T} \sum_{t=1}^{T} x_t x_t^* \quad .$$

The autocorrelation function value r(0) is equal to P_1.

Once the N-point-long prediction error filter outputs and coefficients are known, the (N+1)-point-long prediction error filter is formed from the N-point-long prediction error filter through the determination of the single coefficient a_{N+1}^{N+1} in the (N+1)-point-long filter. With the aid of the Levinson recursion relationship between the N-point-long and (N+1)-point-long prediction

error filters, the forward and backward (N+1) -point-long filter outputs can be expressed as a linear combination of the N-point-long filter outputs:

$$q_{t+N}^{N+1} = a_{N+1}^{N+1} \; p_t^N + q_{t+N}^N$$

$$p_t^{N+1} = p_t^N + \left(a_{N+1}^{N+1}\right)^* \; q_{t+N}^N \; ,$$

where q_t^N and p_t^N are the outputs at time t from the N-point-long forward and backward prediction error filters, respectively. Figure III-1 illustrates this relationship, which was first stated explicitly in the literature by Andersen (Andersen, 1974). The coefficient a_{N+1}^{N+1} is chosen to minimize the average

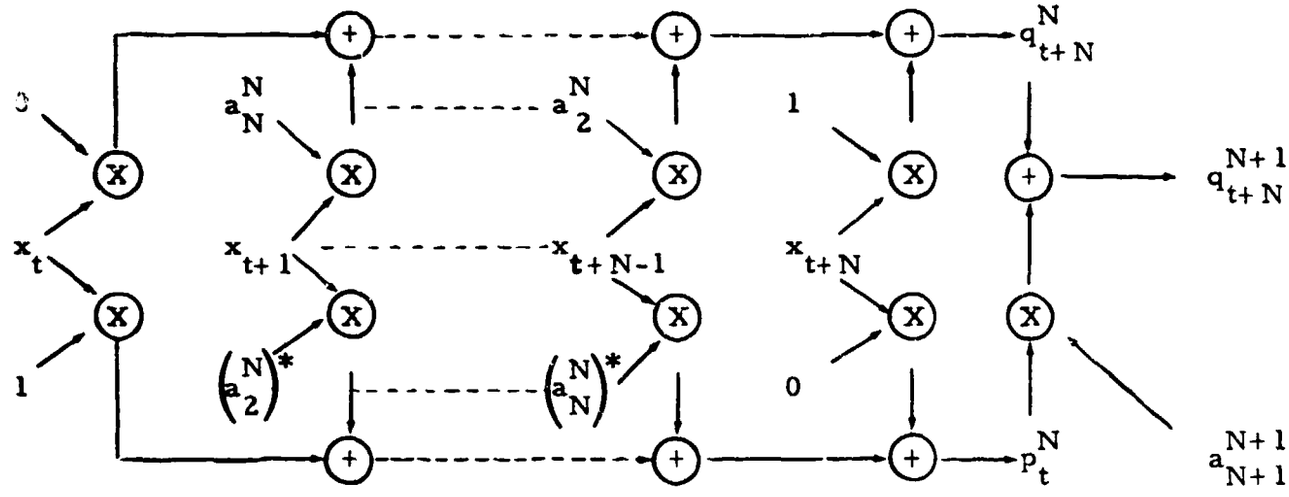
$$E_{N+1} = \frac{1}{2(T-N)} \sum_{t=1}^{T-N} \left[\left| q_{t+N}^N + a_{N+1}^{N+1} \; p_t^N \right|^2 + \left| p_t^N \left(a_{N+1}^{N+1}\right)^* q_{t+N}^N \right|^2 \right]$$

$$= \frac{1}{2(T-N)} \left\{ \left[\sum_{t=1}^{T-N} p_t^N \left(p_t^N\right)^* + q_{t+N}^N \left(q_{t+N}^N\right)^* \right] \left[1 + a_{N+1}^{N+1} \left(a_{N+1}^{N+1}\right)^* \right] \right.$$

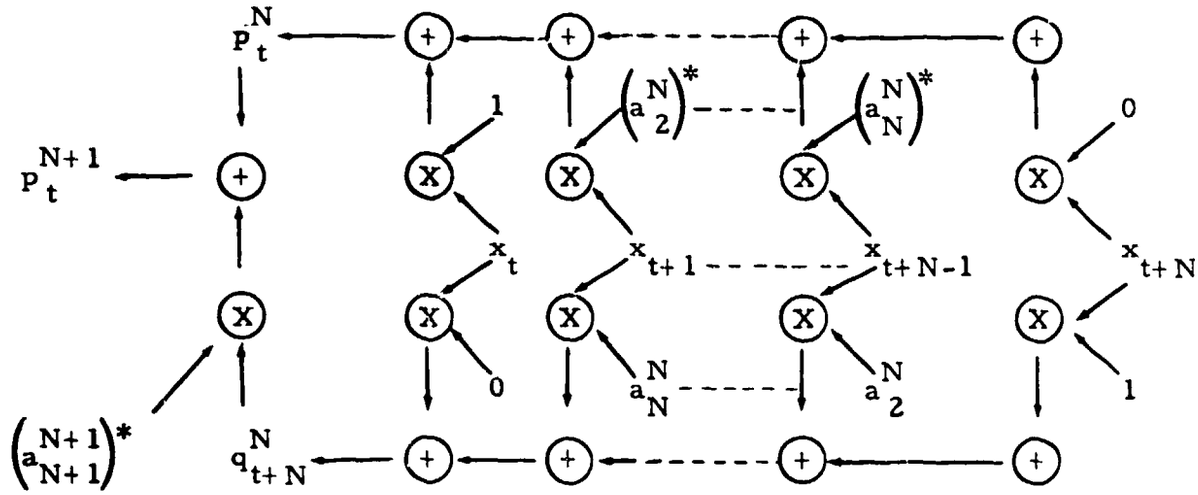
$$\left. + 2 \left[\sum_{t=1}^{T-N} p_t^N \left(q_{t+N}^N\right)^* \right] a_{N+1}^{N+1} + 2 \left[\sum_{t=1}^{T-N} \left(p_t^N\right)^* q_{t+N}^N \right] \left(a_{N+1}^{N+1}\right)^* \right\}$$

of the forward and backward prediction error powers. E_{N+1} is a minimum when its partial derivatives with respect to the real and imaginary parts of a_{N+1}^{N+1} are zero:

$$0 = \frac{\partial E_{N+1}}{\partial \mathrm{Re}\; a_{N+1}^{N+1}} + i \; \frac{\partial E_{N+1}}{\partial \mathrm{Im}\; a_{N+1}^{N+1}}$$



(a) **Forward Prediction Error Filter Output:** $q_{t+N}^{N+1} = a_{N+1}^{N+1}\, p_t^N + q_{t+N}^N$



(b) **Backward Prediction Error Filter Output:** $p_t^{N+1} = p_t^N + \left(a_{N+1}^{N+1}\right)^* q_{t+N}^N$

FIGURE III-1

FORMATION OF THE (N+1)-POINT-LONG PREDICTION ERROR FILTER OUTPUTS FROM THE N-POINT-LONG PREDICTION ERROR FILTER OUTPUTS USING THE LEVINSON RECURSION RELATIONSHIP

$$= \frac{1}{(T-N)} \left\{ \left[\sum_{t=1}^{T-N} p_t^N \left(p_t^N\right)^* + q_{t+N}^N \left(q_{t+N}^N\right)^* \right] a_{N+1}^{N+1} + 2\left[\sum_{t=1}^{T-N} \left(p_t^N\right)^* q_{t+N}^N \right] \right\},$$

so that a_{N+1}^{N+1} is chosen to be

$$a_{N+1}^{N+1} = - \frac{2 \sum_{t=1}^{T-N} \left(p_t^N\right)^* q_{t+N}^N}{\sum_{t=1}^{T-N} \left[p_t^N \left(p_t^N\right)^* + q_{t+N}^N \left(q_{t+N}^N\right)^* \right]} .$$

Since the summation in the numerator is the inner product between the vectors $(p_1, p_2, \ldots, p_{T-N})$ and $(q_{N+1}, q_{N+2}, \ldots, q_t)$ and the denominator is the sum of the squared magnitudes of these two vectors, the absolute value of a_{N+1}^{N+1} never exceeds one. If the absolute values of $a_2^2, a_3^3, \ldots, a_{N+1}^{N+1}$ are all less than one, then the forward prediction error is minimum phase (see Appendix A). Substitution of a_{N+1}^{N+1} into the expression for E_{N+1} yields

$$E_{N+1} = \frac{1}{2(T-N)} \left[\sum_{t=1}^{T-N} p_t^N \left(p_t^N\right)^* + q_{t+N}^N \left(q_{t+N}^N\right)^* \right] \left[1 - a_{N+1}^{N+1} \left(a_{N+1}^{N+1}\right)^* \right] .$$

The (N+1)-point-long forward prediction error filter is determined from the Levinson recursion relationship:

$$\begin{bmatrix} 1 \\ a_2^{N+1} \\ \vdots \\ a_N^{N+1} \\ a_{N+1}^{N+1} \end{bmatrix} = \begin{bmatrix} 1 \\ a_2^N \\ \vdots \\ a_N^N \\ 0 \end{bmatrix} + a_{N+1}^{N+1} \begin{bmatrix} 0 \\ \left(a_N^N\right)^* \\ \vdots \\ \left(a_2^N\right)^* \\ 1 \end{bmatrix} .$$

The prediction error power for the (N+1)-point-long filter is also determined from the Levinson recursive relationship:

$$P_{N+1} = P_N \left[1 - a_{N+1}^{N+1} \left(a_{N+1}^{N+1} \right)^* \right] .$$

This power is, in general, not equal to E_{N+1}. If the absolute values of a_2^2, a_3^3, ... , a_{N+1}^{N+1} are all less than one, then the matrix R_{N+1} is positive definite since the successive determinants

$$|R_k| = \prod_{j=1}^{k} P_j \qquad (k = 1, 2, \ldots , N+1)$$

are all positive. If desired, the corner elements r(N) and $r^*(N)$ in the matrix R_{N+1} can be recursively determined from the bottom row of the (N+1)-point-long prediction error filter design matrix equation:

$$r(N) = -\sum_{j=1}^{N} \left(a_{j+1}^{N+1} \right)^* r(N-j).$$

Thus the matrix R_{N+1} can be determined from P_1 and the successive prediction error filters.

From the prediction error power P_N and the N-point-long prediction error filter coefficients, the maximum entropy spectrum

$$P(f) = \frac{P_N \Delta t}{\left| 1 + \sum_{j=1}^{N=1} a_{j+1}^{N} \, e^{i2\pi f j \Delta t} \right|^2}$$

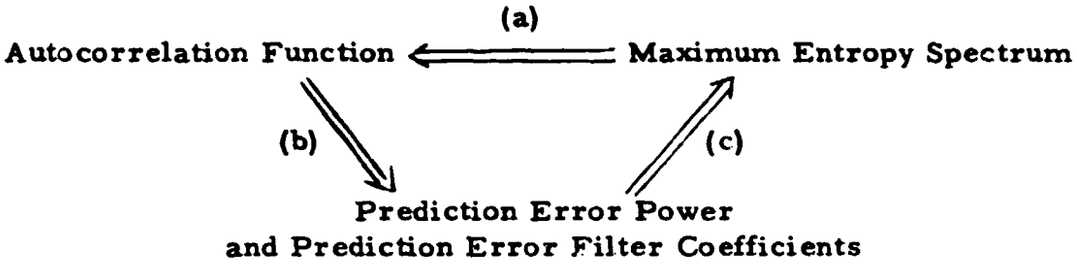
corresponding to the Burg technique can be calculated. Empirical comparisons of the maximum entropy spectrum using the Burg technique with the maximum entropy spectrum using estimated autocorrelation functions (e. g., Radoski, Fougere, and Zawalick, 1974) indicate greater resolution and spectral accuracy is possible with the Burg technique. As far as known, however, there are no definitive statistical studies to support this claim. Furthermore, there is no known objective method for determining the order of the autoregression model when the Burg technique is used. Akaike's criteria appear to be inextricably linked to autocorrelation functions estimated by the formula

$$r(j) = \frac{1}{T-j} \sum_{t=1}^{T-j} x_t \, x^*_{t+j} \quad .$$

The statistical evaluation of maximum entropy spectra using the Burg technique, therefore, appears to be an area meriting more detailed scrutiny. At the present time, the choice of the prediction error filter length for the Burg technique is a matter of subjective judgment.

Since the Burg technique computes the prediction error filter coefficients and filter output power directly, the autocorrelation function is not needed for the maximum entropy spectrum. In fact, the autocorrelation function lags 0 to N-1 or the order N-1 maximum entropy spectrum or the N-point long prediction error filter and its output power contain equivalent information. Figure III-2 illustrates the relationships between these three functions. A knowledge of one permits the other two to be determined. Figure III-2 presents some, but by no means all, of the ways to accomplish the transformations between these three functions.

The other principal use for autocorrelation functions is in time-domain digital filter design, where the inverse of the autocorrelation matrix R_N is needed. As shown earlier in this paper, however, the inverse is



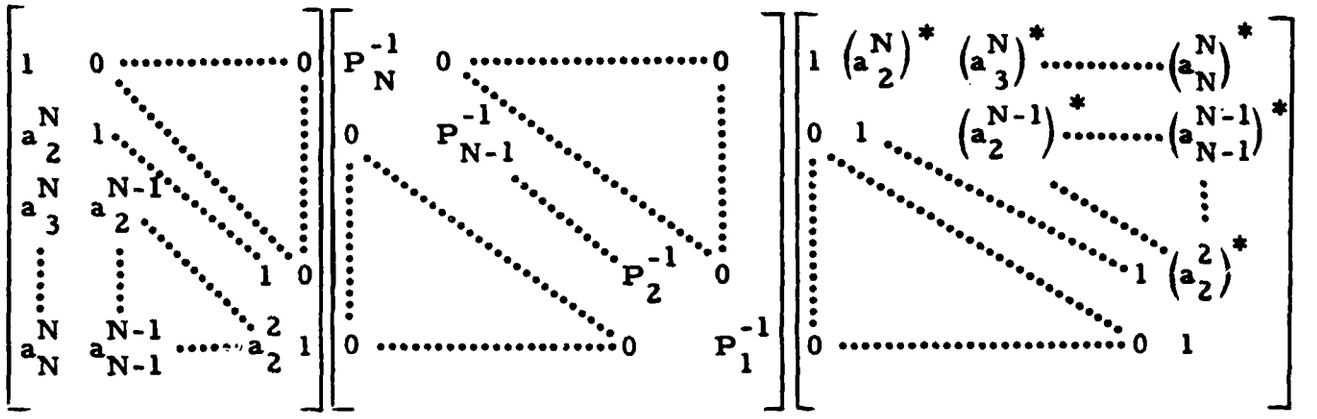
$$(a)\quad r(j) = \int_{-W/2}^{W/2} P(f)\, e^{i2\pi f j \Delta t}\, df$$

$$(b)\quad \begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_2^N \\ \vdots \\ a_N^N \end{bmatrix} = \begin{bmatrix} P_N \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$(c)\quad P(f) = \frac{P_N \Delta t}{\left| 1 + \sum_{j=1}^{N-1} a_{j+1}^N\, e^{i2\pi f j \Delta t} \right|^2}$$

FIGURE III-2

RELATIONSHIP BETWEEN AUTOCORRELATION FUNCTION, PREDICTION ERROR FILTER AND MAXIMUM ENTROPY SPECTRUM



which can be obtained from the successive prediction error filters and their error powers.

SECTION IV
THE K-LINE SPECTRUM

When the crosspower spectrum matrix for an equally-spaced line array is available, there is a wavenumber analogue to the frequency-domain maximum entropy spectrum described previously in this paper. It is known as the K-line spectrum and was developed by John Parker Burg while at Texas Instruments.

If a space-time wavefield can be described as a superposition of plane waves, then the inverse Fourier transform

$$r(j) = \int_{-W/2}^{W/2} P(k)\, e^{-i2\pi kj\Delta x}\, dk$$

of the wavenumber spectrum P(k) is equal to the crosspower spectrum for two sensors at a spatial displacement of $j\Delta x$, where Δx is the distance between two successive sensors in the line array. This fact requires that all crosspower spectra corresponding to the same spatial displacement be identical:

$$r(j) = \Phi_{m,\,m+j}(f) = \int_{-\infty}^{\infty} \phi_{m,\,m+j}(\tau)\, e^{-i2\pi f\tau}\, d\tau\,,$$

where

$$\phi_{m,\,m+j}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g_m(t)\; g_{m+j}(t-\tau)\; dt$$

is the crosscorrelation function at a time lag τ between the output $g_m(t)$ of the m-th sensor and the output $g_{m+j}(t)$ of the (m+j)-th sensor, and where t denotes the time of the sensor output. As a result, the NXN crosspower spectrum matrix

$$R_N = \begin{bmatrix} \Phi_{11}(f) & \Phi_{12}(f) \cdots \Phi_{1N}(f) \\ \Phi_{21}(f) & \Phi_{22}(f) \cdots \Phi_{2N}(f) \\ \vdots & \vdots \\ \Phi_{N1}(f) & \Phi_{N2}(f) \cdots \Phi_{NN}(f) \end{bmatrix} = \begin{bmatrix} r(0) & r(1) \cdots r(N-1) \\ r^*(1) & r(0) \cdots r(N-2) \\ \vdots & \vdots \\ r^*(N-1) & r^*(N-2) \cdots r(0) \end{bmatrix}$$

for an N-element array with sensor coordinates 0, Δx, $2\Delta x$ ... , $(N-1)\Delta x$ is a Toplitz matrix as well as a Hermitian non-negative definite matrix.

If $W = 1/\Delta x$, $z = e^{i2\pi k/W}$, and the vectors V and A are, respectively,

$$V = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{N-1} \end{bmatrix} = \begin{bmatrix} 1 \\ e^{i2\pi k \Delta x} \\ \vdots \\ e^{i2\pi k(N-1)\Delta x} \end{bmatrix} \text{ and } A = \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix},$$

the wavenumber maximum entropy spectrum is

$$P(\kappa) = \frac{1}{W} \frac{P_N}{V^{H} A A^{H} V} = \frac{P_N \Delta x}{\left| 1 + \sum_{j=1}^{N-1} a_{j+1} e^{-i2\pi kj\Delta x} \right|^2},$$

where P_N and $a_2, a_3, \ldots, a_N$ are solutions of the matrix equation

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} P_N \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

The proof is very similar to the proof for the frequency-domain maximum entropy spectrum and has been given previously (Barnard, 1969).

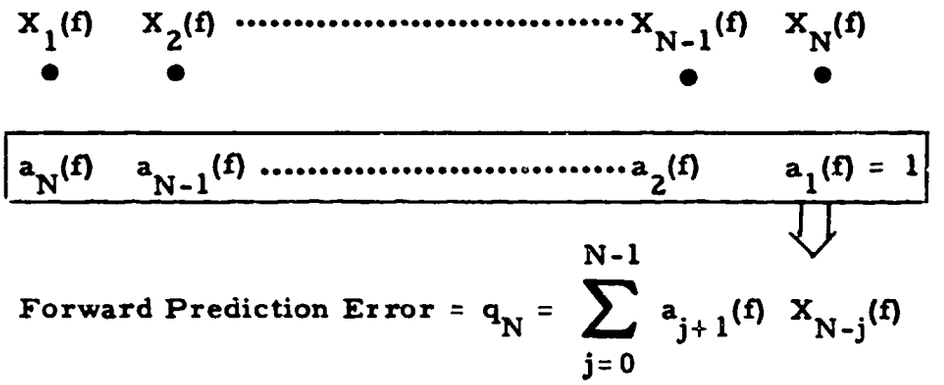
In the case of the wavenumber maximum entropy spectrum, the forward prediction error filter with weights $(1, a_2, \ldots, a_N)$ is a spatial filter whose output is the error in estimating the N-th sensor from the first N-1 sensors. Figure IV-1 illustrates this situation. In the figure, $X_j(f)$ denotes the Fourier transform of the j-th sensor output. The backward prediction error filter with weights $(a_N^*, \ldots, a_2^*, 1)$ outputs the error in estimating the first sensor from the second through N-th sensors. The term P_N in the filter design matrix equation is the spatial prediction error power density spectrum and is equal to both $\overline{q_N q_N^*}$ and $\overline{p_1 p_1^*}$, where the vinculum denotes the crosspower spectrum (or autopower spectrum) of the two quantities below it.

In practice, the assumption of space-stationarity is not satisified and the elements along any diagonal are not equal:

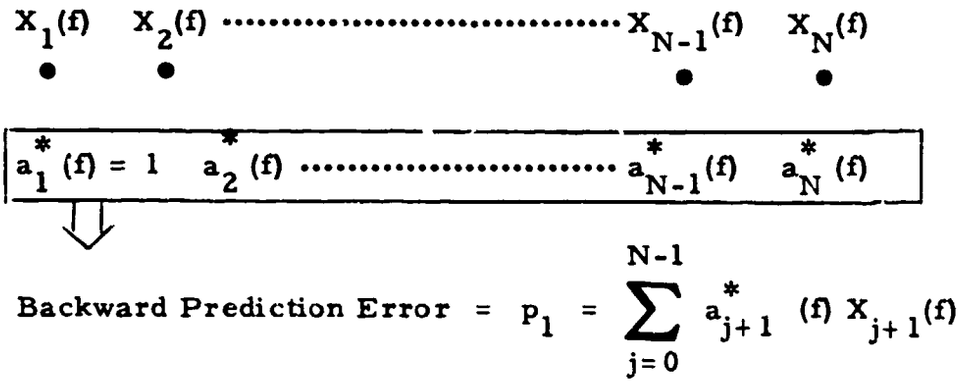
$$\Phi_{m,\,m+j}(f) \neq \Phi_{n,\,n+j}(f) \qquad (m \neq n).$$

The easiest way to remedy this problem is to average the elements along each diagonal of the crosspower spectrum matrix. The ensuing spectrum is the wavenumber analogue of the standard autoregressive spectrum. If this procedure is performed, however, the resulting Toeplitz matrix may not be non-negative definite.

Another way to compute the wavenumber maximum entropy spectrum is to use a modified version of the Burg technique. The procedure begins by averaging the autopower spectra to estimate the 1-point-long prediction error power density spectrum:



(a) Forward Spatial Prediction Error Filter



(b) Backward Spatial Prediction Error Filter

FIGURE IV-1

SPATIAL PREDICTION ERROR FILTERING

$$P_1 = \frac{1}{N} \sum_{m=1}^{N} \overline{X_m X_m^*} = \frac{1}{N} \sum_{m=1}^{N} \Phi_{mm}(f) \quad .$$

Then the algorithm continues recursively with the minimization of the spatial prediction error power density spectrum

$$E_{M+1} = \frac{1}{2(N-M)} \sum_{m=1}^{N-M} \left[\overline{\left| q_{m+M}^{M} + a_{M+1}^{M+1} \; p_m^M \right|^2} + \overline{\left| p_m^M + \left(a_{M+1}^{M+1}\right)^* q_{m+M}^M \right|^2} \right]$$

for the $(M+1)$-point-long filter by setting a_{M+1}^{M+1} equal to

$$a_{M+1}^{M+1} = \frac{-2 \sum_{m=1}^{N-M} \overline{\left(p_m^M\right)^* \; q_{m+M}^M}}{\sum_{m=1}^{N-M} \left[\overline{p_m^M \left(p_m^M\right)^*} + \overline{q_{m+M}^M \left(q_{m+M}^M\right)^*} \right]} = \frac{-2S_1}{S_2} \quad ,$$

where S_1 and S_2 are

$$S_1 = \sum_{m=1}^{N-M} \overline{\left[a_M^M \cdots\cdots a_2^M \;\; 1\right] \begin{bmatrix} X_{m+1} \\ \vdots \\ X_{m+M-1} \\ X_{m+M} \end{bmatrix} \left[X_m^* \;\; X_{m+1}^* \cdots\cdots X_{m+M-1}^*\right] \begin{bmatrix} 1 \\ a_2^M \\ \vdots \\ a_M^M \end{bmatrix}}$$

$$\left[a_M^M \cdots\cdots a_2^M \quad 1\right]\left\{\sum_{m=1}^{N-M}\begin{bmatrix} \phi_{m+1,m}^{(f)} & \phi_{m+1,m+1}^{(f)} & \cdots\cdots & \phi_{m+1,m+M-1}^{(f)} \\ \vdots & \vdots & & \vdots \\ \phi_{m+M-1,m}^{(f)} & \phi_{m+M-1,m+1}^{(f)} & \cdots & \phi_{m+M-1,m+M-1}^{(f)} \\ \phi_{m+M,m}^{(f)} & \phi_{m+M,m+1}^{(f)} & \cdots\cdots & \phi_{m+M,m+M-1}^{(f)} \end{bmatrix}\right\}\begin{bmatrix} 1 \\ a_2^M \\ \vdots \\ a_M^M \end{bmatrix}$$

$$= S_2 = \sum_{m=1}^{N-M}\left\{\left[1 \quad \left(a_2^M\right)^* \cdots\cdots \left(a_M^M\right)^*\right]\begin{bmatrix} X_m \\ X_{m+1} \\ \vdots \\ X_{m+M-1}^* \end{bmatrix}\overline{\left[X_m \quad X_{m+1}^* \cdots\cdots X_{m+M-1}^*\right]}\begin{bmatrix} 1 \\ a_2^M \\ \vdots \\ a_M^M \end{bmatrix}\right.$$

$$\left. + \left[1 \quad \left(a_2^M\right)^* \cdots\cdots \left(a_M^M\right)^*\right]\begin{bmatrix} X_{m+M}^* \\ X_{m+M-1}^* \\ \vdots \\ X_{m+1}^* \end{bmatrix}\overline{\left[X_{m+M} \quad X_{m+M-1} \cdots\cdots X_{m+1}\right]}\begin{bmatrix} 1 \\ a_2^M \\ \vdots \\ a_M^M \end{bmatrix}\right\}$$

$$= \left[1 \quad \left(a_2^M\right)^* \cdots\cdots \left(a_M^M\right)^*\right]\begin{bmatrix} & & \\ & C & \\ & & \end{bmatrix}\begin{bmatrix} 1 \\ a_2^M \\ \vdots \\ a_M^M \end{bmatrix},$$

and where

$$[C] = \sum_{m=1}^{N-M} \begin{bmatrix} \Phi_{mm}(f) & \Phi_{m+M,m+M}(f) & \Phi_{m,m+1}(f) & \Phi_{m+M-1,m+M}(f) \cdots\cdots & \Phi_{m,m+M-1}(f) & \Phi_{m+1,m+M}(f) \\ \Phi_{m+1,m}(f) & \Phi_{m+M,m+M-1}(f) & \Phi_{m+1,m+1}(f) & \Phi_{m+M-1,m+M-1}(f) \cdots\cdots \Phi_{m+1,m+M-1}(f) & \Phi_{m+1,m+M-1}(f) & \\ \Phi_{m+M-1,m}(f) & \Phi_{m+M,m+1}(f) & \Phi_{m+M-1,m+1}(f) & \Phi_{m+M-1,m+1}(f) \cdots\cdots \Phi_{m+M-1,m+M-1}(f) & \Phi_{m+1,m+1}(f) & \end{bmatrix} .$$

Aside from the way a_{M+1}^{M+1} is computed, the rest is the same as for the Burg technique proper. Normally (but not always) the procedure continues until the prediction error filter length is the same as the number of sensors.

The absolute value of a_{M+1}^{M+1} never exceeds one since

$$\left| \overline{\left(p_M^M\right)^* \, q_{m+M}^M} \right| \leq \frac{1}{2} \left[\overline{p_m^M \left(p_m^M\right)^*} + \overline{q_{m+M}^M \left(q_{m+M}^M\right)^*} \right] .$$

(The absolute value of a crosspower spectrum never exceeds the average value of the two corresponding autopower spectra).

Because the prediction error filter outputs are not available, the crosspower spectrum matrix must be used in applying the Burg technique to wavenumber spectra. This fact makes the required computations more cumbersome, especially for arrays with many sensors. If, in certain situations, some loss in spectral resolution and accuracy can be tolerated, the wavenumber analogue to the standard autoregressive spectrum provides a reasonable alternative.

SECTION V
REFERENCES

Aasnaes, Hans Bent, and Thomas Kailath, December 1973, "An Innovations Approach to Least-Square Estimation - Part VII: Some Applications of Vector Autoregressive-Moving Average Models," IEEE Transactions on Automatic Control, Volume AC-18, No. 6, pp. 601-607.

Akaike, Hirotugu, 1969, "Fitting Autoregressions for Prediction," Ann Inst. Statist. Math., Volume 21, pp. 243-247.

Akaike, Hirotugu, 1970a, "Statistical Predictor Identification," Ann. Inst. Statist. Math, Volume 22, pp. 203-217.

Akaike, Hirotugu, 1970b, "On a Semi-Automatic Power Spectrum Estimation Procedure," in Proc. 3rd Hawaii Int. Conf. System Sciences (part 2), pp. 974-977.

Akaike, Hirotugu, 1971a, "Autoregressive Model Fitting for Control," Ann. Inst. Statist. Math, Volume 23, pp. 163-180.

Akaike, Hirotugu, 1971b, "Information Theory and an Extension of the Maximum Likelihood Principle," Inst. Statist. Math. Res. Memo 46, presented at the 2nd Int. Symp. Information Theory, Tsahkadsor, Armenian SSR, Sept. 2-8, 1971; also Supp. to Problems of Control and Information Theory, Akademial Kiado, Budapest, 1974, pp. 267-281.

Akaike, Hirotugu, 1972a, "Use of an Information Theoretic Quantity for Statistical Model Identification," in Proc. 5th Hawaii Int. Conf. System Sciences, pp. 249-250.

Akaike, Hirotugu, 1972b, "Automatic Data Structure Search by the Maximum Likelihood," in Computers in Biomedicine (supplement to Proc. 5th Hawaii Int. Conf. on System Sciences), pp. 99-101. North Hollywood, California; Western Periodicals.

Akaike, Hirotugu, December 1974, "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, Volume AC-19, No. 6, pp. 716-723.

Anderson, Nils O., February 1974, "On the Calculation of Filter Coefficients for Maximum Entropy Spectral Analysis," Geophysics, Volume 39, No. 1, pp. 69-72.

Barnard, Thomas E., 14 May 1969, "Analytical Studies of Techniques for the Computation of High-Resolution Wavenumber Spectra," Advanced Array Research Special Report No. 9, Texas Instruments Incorporated.

Burg, John Parker, 31 October 1967, "Maximum Entropy Spectral Analysis," presented at the 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma.

Burg, John Parker, 1968, "A New Analysis Technique for Time Series Data," presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, Netherlands.

Burg, John Parker, 30 April 1974, "Report on Project to Research Develop and Implement On-Line Shipboard Deconvolution, Stacking and Digital Recording of Seismic Reflection Profile Data," Time and Space Processing, Inc., Palo Alto, California.

Chen, W. Y., and G. R. Stegun, 10 July 1974, "Experiments with Maximum Entropy Power Spectra of Sinusoids," Journal of Geophysical Research, Volume 79, No. 20, pp. 3019-3022.

Durbin, J., 1960, "The Fitting of Time-Series Models," Rev. Intern. Statist. Volume 28, pp. 233-244.

Frost, Paul A., and Thomas Kailath, June 1971, "An Innovations Approach to Least-Squares Estimation - Part III: Nonlinear Estimation in White Gaussian Noise," IEEE Transactions on Automatic Control, Volume AC-16, No. 3,pp. 217-226.

Gersh, Will, and David R. Sharpe, August 1973, "Estimation of Power Spectra with Finite-Order Autoregressive Models," IEEE Transactions on Automatic Control, Volume AC-18, No. 4, pp. 367-369.

Gevers, Michel R., and Thomas Kailath, December 1973, "An Innovations Approach to Least-Squares Estimation - Part VI: Discrete - Time Innovations Representations and Recursive Estimation," IEEE Transactions on Automatic Control, Volume AC-18 No. 6, pp. 588-600.

Jones, Richard H., December 1974, "Identification and Autoregressive Spectrum Estimation," IEEE Transactions on Automatic Control, Volume AC-19, No. 6, pp. 894-898.

Jury, E. I., 1962, "A Simplified Stability Criteria for Linear Discrete Systems," Proceedings of the IRE, Volume 50, pp. 1493-1500.

Kailath, Thomas, December 1968, "An Innovations Approach to Least-Squares Estimation, Part I: Linear Filtering in Additive White Noise," IEEE Transactions on Automatic Control, Volume AC-13, No. 6, pp. 646-655.

Kailath, Thomas, and Paul A. Frost, December 1968, "An Innovations Approach to Least-Squares Estimation, Part II: Linear Smoothing in Additive White Noise, IEEE Transactions on Automatic Control, Volume AC-13, No. 6, pp. 655-660.

Kailath, Thomas, May 1970, "The Innovations Approach to Detection and Estimation Theory," Proceedings of the IEEE, Volume 58, No. 5, pp. 680-695.

Kailath, Thomas, March 1974, "A View of Three Decades of Linear Filtering Theory," IEEE Transactions on Information Theory, Volume IT-20, No. 2, pp. 146-181.

Kailath, Thomas, and Roger A. Geesey, December 1971, "An Innovations Approach to Least Squares Estimation - Part IV: Recursive Estimation Given Lumped Covariance Functions," IEEE Transactions on Automatic Control, Volume AC-16, No. 6, pp. 720-727.

Kailath, Thomas, and Roger A. Geesey, October 1973, "An Innovations Approach to Least-Squares Estimation - Part V: Innovations Representations and Recursive Estimation in Colored Noise," IEEE Transactions on Automatic Control, Volume AC-18, No. 5, pp. 435-453.

King, W. R., W. H. Swindell, and L. J. O'Brien, 28 February 1974, "Final Report on Development of a Curvilinear Ray Theory Model and Maximum Entropy Spectral Analysis," Texas Instruments Incorporated.

Kromer, R. E., 1969, "Asymptotic Properties of the Autoregressive Spectral Estimator, " Ph. D. dissertation, Dept. State., Stanford Univ., Stanford, Calif., Tech. Rep. 13.

Lacoss, Richard T., August 1971, "Data Adaptive Spectral Analysis Methods," Geophysics, Volume 36, No. 4, pp. 661-675.

Levinson, Norman, January 1947, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction," Journal of Math. and Physics, Volume 25, pp. 261-278; also available as Appendix B in Norbert Wiener, Extrapolation, Interpolation, and Smoothing Stationary Time Series with Engineering Applications, New York, Technology Press and Wiley, 1949.

McDonough, R. N., December 1974, "Maximum-Entropy Spatial Processing of Array Data," Geophysics, Volume 39, No. 6, pp. 843-851.

Marden, M., 1949, The Geometry of the Zeros of a Polynomial in a Complex Variable, Math. Surveys, No. 3, American Mathematical Society, New York.

Radoski, Henry R., Paul F. Fougere, and Edward J. Zawalick, 13 February 1974, "A High Resolution Power Spectral Estimate: The Maximum Entropy Method," Air Force Surveys in Geophysics, No. 282, Space Physics Laboratory, Air Force Cambridge Research Laboratories, Air Force Systems Command, Report No. AFCRL-TR-74-0088.

Riley, Don C., and John Parker Burg, 1972, "Time and Space Adaptive Deconvolution Filters," presented at the 1972 Convention of the Society of Exploration Geophysicists, Anaheim, California.

Ulrych, Tad J., 10 March 1972, "Maximum Entropy Power Spectrum of Truncated Sinusoids," Journal of Geophysical Research, Volume 77, No. 8, pp. 1396-1400.

Ulrych, Tad J., D. E. Smylie, O. G. Jensen, and G. K. C. Clarke, 10 August 1973, "Predictive Filtering and Smoothing of Short Records by Using Maximum Entropy," Journal of Geophysical Research, Volume 78, No. 23, pp. 4959-4964.

APPENDIX A

TEST FOR MINIMUM PHASE SAMPLE POINT FILTERS

(by John Parker Burg)

In this appendix, a simple test is developed to determine whether a finite length sample point operator is minimum phase. The development of this test, which is known as Jury's stability test (Jury, 1962) in the theory of sampled data control systems, is contained in the book The Geometry of the Zeros by M. Marden (Marden, 1949). The justification for presenting this test and two important theorems about the zeroes of a polynominal in condensed form is their apparent newness to geophysical workers in the field of digital filter theory.

A sample point operator is defined here to be minimum phase if and only if both the filter and its inverse sample point filter are

1) physically realizable (i.e., their z-tranforms are Taylor series in z of the form $a_o + a_1 z + \ldots + a_n z^n + \ldots$), and

2) stable (i.e., their z-transform series converge for $|z| = 1$).

This definition is equivalent to saying that the z-transform of a minimum phase sample point operator is both analytic and non-zero on and inside the unit circle. Since a finite length, physically realizable, sample point operator is analytic for $|z| < \infty$, the test for minimum phase becomes one of determining whether all the zeroes of a polynominal, $a_o + a_1 z + \ldots + a_N z^N$, are outside the unit circle.

To develop the test, two important theorems about the zeroes of a polynominal are first proved. These two theorems and the test are valid for polynominals with complex coefficients.

Principle of Argument Theorem

Let C be a simple closed Jordan curve in the complex plane and let F(z) be a polynominal in z, none of whose zeroes lie on C. Let Δ be the total phase shift in F(z) as the point z traverses C once in a counterclockwise direction. Then the number of zeroes of F(z) inside C, counted with their multiplicities, is given by $\Delta/2\pi$.

Proof: Factoring $F(z) = a_o + a_1 z + \ldots + a_N z^N$ into $a_N \prod_{i=1}^{N} (z-\alpha_i)$, we see that each root which lies inside C contributes 2π to the total phase shift of F(z), but that each root outside of C has zero contribution.

Rouché's Theorem

If P(z) and Q(z) are two polynominals in z for which $|P(z)| > |Q(z)|$ on a simple closed Jordan curve C, then the polynomial F(z) = P(z) + Q(z) has the same number of zeroes inside C as does P(z).

Proof: We should first note that since $|P(z)| > |Q(z)|$ on C, P(z) and F(z) cannot have any zeroes on C. Writing $F(z) = P(z)\,[1 + Q(z)/P(z)]$, the total change in the argument of F(z), as C is traversed once in a counterclockwise direction, is the sum of the total phase shift of P(z) and $1 + Q(z)/P(z)$. But since $|Q(z)| < |P(z)|$ for z on C, the real part of $1 + Q(z)/P(z)$ is always positive for z on C and thus has a total phase shift of zero. Therefore, F(z) and P(z) have the same total phase shift and thus, from the Principle of Argument Theorem, they have the same number of zeroes inside C.

Minimum Phase Theorem

A finite length sample point operator is minimum phase if and only if its z-transform is given by $a_o Q_N(z)$, where $Q_N(z)$ satisfies the recursive procedure

$$Q_o(z) = 1$$

$$Q_n(z) = Q_{n-1}(z) + r_n z^n \left[Q_{n-1}\left(z^{-1*}\right)\right]^* \qquad \text{(A.1)}$$

with all $|r_n| < 1$.

Proof: We prove that a filter generated by this recursive procedure will be minimum phase by noting that

1) The $Q_n(z)$ are of the form $1 + a_1 z + a_2 z^2 + \ldots + a_n z^n$ and thus are analytic on and inside the unit circle, and

2) Starting with $Q_o(z) = 1$, which has no zeroes on or inside the unit circle, we see that $\left[Q_{n-1}(z)\right]^* = \left[Q_{n-1}\left(z^{-1*}\right)\right]^*$ for $|z| = 1$ and thus $\left|Q_{n-1}(z)\right| > \left|r_n z^n \left[Q_{n-1}\left(z^{-1*}\right)\right]^*\right|$ on the unit circle. Therefore, using Rouché's Theorem repeatedly, $Q_N(z)$ will have no zeroes on or inside the unit circle and thus $a_o Q_N(z)$ will be minimum phase.

That any finite length minimum phase filter can be obtained from (A.1) is proved by seeing that the reverse of the recursive procedure is unique and that all $|r_n|$ will be less than one. Letting $F_N(z)$ be the z-transform of an N+1 point minimum phase filter after normalizing the first term to one, we can write

$$\begin{aligned} F(z) &= 1 + b_1 z + b_2 z^2 + \ldots + b_{N-1} z^{N-1} + b_N z^N \\ &= \left[1 + c_1 z + c_2 z^2 + \ldots + c_{N-1} z^{N-1}\right] \\ &\quad + b_n \left[c^*_{N-1} z + c^*_{N-2} z^2 + \ldots + c^*_1 z^{N-1} + z^N\right], \qquad \text{(A.2)} \end{aligned}$$

where the c_s are determined by the equation

$$\begin{bmatrix} 1 & b_N \\ b_N^* & 1 \end{bmatrix} \begin{bmatrix} c_s \\ c_{N-s}^* \end{bmatrix} = \begin{bmatrix} b_s \\ b_{N-s}^* \end{bmatrix} . \qquad \text{(A. 3)}$$

Since $F_N(z)$ is minimum phase, all of its roots, α_i, are greater than one in magnitude. Therefore, since

$$b_N = \prod_{i=1}^{N} (-\alpha_i)^{-1} ,$$

$|b_N| <$ | and (A. 3) will always have a unique solution. Furthermore, from Rouché's Theorem, the first polynominal on the right hand side of (A. 2) will also be minimum phase. Thus, the recursive procedure of (A. 1) is uniquely reversible with all $|r_n| < 1$ for a finite length minimum phase filter.

APPENDIX B

COMPARISONS OF THE CHARACTERISTICS OF MAXIMUM ENTROPY AND DISCRETE FOURIER TRANSFORM SPECTRAL ESTIMATION

In this appendix, a brief comparison is made between the characteristics of maximum entropy and discrete Fourier transform spectral analysis. This appendix is taken from a contractor report (King, Swindell, and O'Brien, 1974) and was written by William H. Swindell, Jr. Several of these comparisons were obtained from an excellent paper by Lacoss (1971).

1. Estimation of Autocorrelation Function

   ME : Uses known lag values unmodified. Unknown lags are estimated in an optimum manner.

   DFT : Weights all lag values with some function introducing spectral window effects in the spectrum.

2. Spectral Window Effects

   ME : No spectral window effects as such are introduced since the autocorrelation function is known or estimated for all lags. However, similar but greatly reduced effects occur when the prediction error filter does not create a perfectly white output.

   DFT : Window effects are always present. Spectral estimates are exactly equal to the convolution of the maximum entropy spectrum with the frequency window corresponding to the particular tapering method employed. These estimates tend to be the convolution of the window function and the true spectrum.

3. Estimate of Peak Power Density of a Pure Tone of Power P

ME : Proportional to P^2N^2 where N is the number of measured autocorrelation function lag values.

DFT: : Proportional to P but subject to picket fence effect from spectral window.

4. Estimate of Bandwidth of a Pure Tone

ME : Proportional to $1/(PN^2)$.

DFT : Proportional to $1/N$.

5. Estimate of Spectral Power in a Pure Tone

ME : Proportional to P.

DFT : Proportional to P/N.

6. Estimate of Line Frequency

ME : Difficult to define but can be estimated very closely. (See Chen and Stegun, 1974, and Ulrych, 1972).

DFT : $\pm$ $1/(2N\Delta t)$.

7. Spectral Reliability

ME : Difficult to define. Asymptotically, for data with low spectral contrast, the degrees of freedom, K, is less than or equal to L/N where L is the number of data points in sample .

DFT : $K \simeq 2L/N$ for Bartlett window.

8. Linearity of Spectra

ME : Estimation is nonlinear. The spectrum of the sum of two time series is not equal to the sum of the spectra of the individual time series.

DFT : Estimation is linear. Superposition of spectra is valid.

9. Resolution of Closely Spaced Spectral Lines

ME : Resolving power is data-dependent and difficult to define but lines can be separated at a frequency increment approximately proportional to $1/N^2$.

DFT : Lines can be separated at a frequency increment proportional to $1/N$.

10. Spectral Line Detectability

The maximum entropy spectrum is clearly superior to the discrete Fourier transform spectrum for weak signals in short data samples where the discrete Fourier transform processing gain is low. For long data samples where the natural line width is greater than $1/(N\Delta t)$, this advantage is reduced. For fixed N, taking longer data gates L increases detectability for the maximum entropy spectrum because of better autocorrelation lag value estimates.

APPENDIX C

THE USE OF THE BURG TECHNIQUE IN FILTERING SHORT RECORDS

This appendix summarizes a method for filtering short data records. The algorithm, described in a recent paper (Ulrych, Smylie, Jensen, and Clarke, 1973), applies a prediction filter to the data to extend the original time series, Fourier transforms the extended time series, multiplies the Fourier transform by the desired filter frequency response, and obtains the filtered time series through inverse Fourier transformation.

Suppose that the Burg technique has been used to design an N-point-long prediction error filter from a short time series x_t (t= 1, 2, ... , T). If the maximum entropy assumption is valid, the autocorrelation function satisfies the relationships incorporated in the following matrix equation:

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r^*(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & & \vdots \\ r^*(N-1) & r^*(N-2) & \cdots & r(0) \\ \hline r^*(N) & r^*(N-1) & \cdots & r^*(1) \\ \vdots & \vdots & & \vdots \\ r^*(N+j) & r^*(N-1+j) & \cdots & r^*(1+j) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} P_N \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix}$$

The optimum (N-1)-point-long forward prediction filter with prediction distance M, on the other hand, is obtained from the filter design equation

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(N-2) \\ r^*(1) & r(0) & \cdots & r(N-3) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(N-2) & r^*(N-3) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} \alpha_1^M \\ \alpha_2^M \\ \vdots \\ \alpha_{N-1}^M \end{bmatrix} = \begin{bmatrix} r^*(M) \\ r^*(M+1) \\ \vdots \\ r^*(M+N-2) \end{bmatrix} .$$

From the first of these two matrix equations, it is clear that

$$\begin{bmatrix} \alpha_1^1 \\ \alpha_2^1 \\ \vdots \\ \alpha_{N-1}^1 \end{bmatrix} = \begin{bmatrix} -a_2 \\ -a_3 \\ \vdots \\ -a_N \end{bmatrix} ,$$

so that the one-step forward prediction filter output at time T+1 is

$$\hat{x}(T+1) = -\sum_{j=1}^{N-1} a_{j+1}\, x(T+1-j).$$

For $2 \leq M \leq N-1$,

$$\begin{bmatrix} r^*(M) \\ r^*(M+1) \\ \vdots \\ r^*(M+N-2) \end{bmatrix} = -a_2 \begin{bmatrix} r^*(M-1) \\ r^*(M) \\ \vdots \\ r^*(M+N-3) \end{bmatrix} - a_3 \begin{bmatrix} r^*(M-2) \\ r^*(M-1) \\ \vdots \\ r^*(M+N-4) \end{bmatrix} - \cdots - a_N \begin{bmatrix} r^*(M-N+1) \\ r^*(M-N+2) \\ \vdots \\ r^*(M-1) \end{bmatrix}$$

$$= -a_2 R_{N-1} \begin{bmatrix} \alpha_1^{M-1} \\ \alpha_2^{M-1} \\ \vdots \\ \alpha_{N-1}^{M-1} \end{bmatrix} - a_3 R_{N-1} \begin{bmatrix} \alpha_1^{M-2} \\ \alpha_2^{M-2} \\ \vdots \\ \alpha_{N-1}^{M-2} \end{bmatrix} - \cdots - a_M R_{N-1} \begin{bmatrix} \alpha_1^1 \\ \alpha_2^1 \\ \vdots \\ \alpha_{N-1}^1 \end{bmatrix}$$

$$-a_{M+1}R_{N-1}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\ -a_{M+2}R_{N-1}\begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix}\ -\cdots-a_{N}R_{N-1}\begin{bmatrix}0\\\vdots\\1\\\vdots\\0\end{bmatrix},$$

and premultiplication of this equation by the inverse of the (N-1)X(N-1) auto-correlation matrix R_{N-1} yields the M-step prediction filter

$$\begin{bmatrix}\alpha_1^M\\ \alpha_2^M\\ \vdots\\ \alpha_{N-1}^M\end{bmatrix} = -a_2\begin{bmatrix}\alpha_1^{M-1}\\ \alpha_2^{M-1}\\ \vdots\\ \alpha_{N-1}^{M-1}\end{bmatrix} -a_3\begin{bmatrix}\alpha_1^{M-2}\\ \alpha_2^{M-2}\\ \vdots\\ \alpha_{N-1}^{M-2}\end{bmatrix} -\cdots-a_M\begin{bmatrix}\alpha_1^{1}\\ \alpha_2^{1}\\ \vdots\\ \alpha_{N-1}^{1}\end{bmatrix}$$

$$-a_{M+1}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\ -a_{M+2}\begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix}\ -\cdots-a_{N}\begin{bmatrix}0\\\vdots\\1\\\vdots\\0\end{bmatrix},$$

so that the M-step forward prediction filter output at time T+M is

$$\hat{x}(T+M) = -\sum_{j=1}^{M-1} a_{j+1}\,\hat{x}(T+M-j) - \sum_{j=M}^{N-1} a_{j+1}\,x(T+M-j).$$

For $M \geq N$,

$$\begin{bmatrix} r^*(M) \\ r^*(M+1) \\ \vdots \\ r^*(M+N-2) \end{bmatrix} = -a_2 R_{N-1} \begin{bmatrix} \alpha_1^{M-1} \\ \alpha_2^{M-1} \\ \vdots \\ \alpha_{M-1}^{M-1} \end{bmatrix} - a_3 R_{N-1} \begin{bmatrix} \alpha_1^{M-2} \\ \alpha_2^{M-2} \\ \vdots \\ \alpha_{N-1}^{M-2} \end{bmatrix} - \cdots - a_N R_{N-1} \begin{bmatrix} \alpha_1^{M+1-N} \\ \alpha_2^{M+1-N} \\ \vdots \\ \alpha_{N-1}^{M+1-N} \end{bmatrix}$$

and

$$\begin{bmatrix} \alpha_1^M \\ \alpha_2^M \\ \vdots \\ \alpha_{N-1}^M \end{bmatrix} = -a_2 \begin{bmatrix} \alpha_1^{M-1} \\ \alpha_2^{M-1} \\ \vdots \\ \alpha_{N-1}^{M-1} \end{bmatrix} - a_3 \begin{bmatrix} \alpha_1^{M-2} \\ \alpha_2^{M-2} \\ \vdots \\ \alpha_{N-1}^{M-2} \end{bmatrix} - \cdots - a_N \begin{bmatrix} \alpha_1^{M+1-N} \\ \alpha_2^{M+1-N} \\ \vdots \\ \alpha_{N-1}^{M+1-N} \end{bmatrix}$$

so that the M-step forward prediction filter output at time $T+M$ is

$$\hat{x}(T+M) = -\sum_{j=1}^{N-1} a_{j+1} \hat{x}(T+M-j).$$

Thus the optimum M-step forward prediction filter outputs $\hat{x}(T+M)$ from the final N-1 measured points can be formed recursively using the (N-1)-point-long prediction filter with the weights $(-a_2, -a_3, \ldots, -a_N)$ obtained from the Burg technique. First, the one-step forward prediction filter output is formed from the final N-1 measured points:

$$\hat{x}(T+1) = -\sum_{j=1}^{N-1} a_{j+1}\, x(T+1-j) .$$

Later, in place of the unmeasured points in the formula

$$\hat{x}(T+M) = -\sum_{j=1}^{N-1} a_{j+1}\, x(T+M-j) ,$$

the prediction filter outputs $\hat{x}(T+M-j)$ are substituted for the points $x(T+M-j)$ where no measurements are available. Similarly, the optimum M-step backward prediction filter outputs $\hat{x}(1-M)$ from the initial N-1 measured points can also be formed recursively starting with

$$\hat{x}(0) = -\sum_{j=1}^{N-1} a_{j+1}^{*}\, x(j)$$

and continuing with the appropriate substitutions in

$$\hat{x}(1-M) = -\sum_{j=1}^{N-1} a_{j+1}^{*}\, x(j+1-M) \qquad (M \geq 2).$$

According to the originators of this method, predicting the time series to a total length four or five times the original length produces acceptable filtering results.

Once the extended time series has been created, the discrete Fourier transform of the extended time series is multiplied by the desired filter response in the frequency domain, and the filtered trace is obtained from the inverse Fourier transform.

For a more detailed exposition of this method and for illustrative exam[illegible], the reader is referred to the original paper (Ulrych, Smylie, Jensen and Clarke, 1973).

APPENDIX D

OBTAINING REFINED ESTIMATES OF SPECTRAL PEAK PARAMETERS

In the maximum entropy spectrum, the power in the band immediately surrounding a spectral peak is a more reliable indication of line strength than the maximum power density. This appendix, written by William H. Swindell, Jr., and based on John Parker Burg's notes, is taken from a contracter report (King, Swindell, and O'Brien, 1974). It presents, among other things, an excellent method for integrating the power around a spectral peak in the maximum entropy spectrum.

Pertinent parameters of a spectral peak in a power density spectrum are:

- Center frequency
- Peak spectral density
- Total spectral power
- Bandwidth at -3 dB points.

Because of the extreme sharpness of some spectral peaks in a maximum entropy power spectrum, there is considerable difficulty in finding the value of the peak density, the total power, and the bandwidth of the peak. Unless the frequency response of the prediction error is measured with a sufficiently fine frequency increment, highly misleading spectral estimates may result. In Figure D-1, an example is shown of a maximum entropy power spectrum which is evaluated at frequencies separated by an increment Δf resulting in power density estimates designated by the large dots. A dashed line connecting the dots is the line which would be seen in an ordinary plot. A peak is indicated

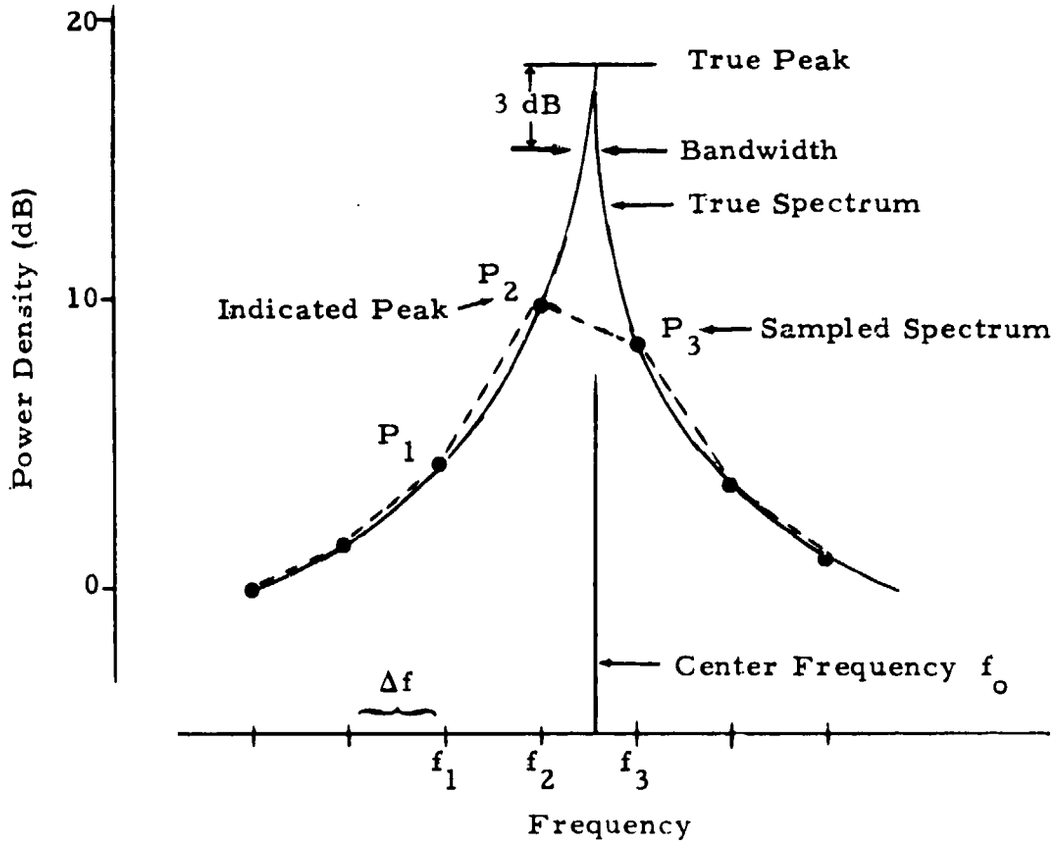


FIGURE D-1

TRUE AND SAMPLED MAXIMUM ENTROPY SPECTRUM

at P_2 with the power densities of P_1 and P_3 on each side. The true response of the filter is shown by the continuous line at a true center frequency of f_o. It is obvious from the figure that the indicated peak density P_2 is almost 10 dB smaller than the true peak and that the center frequency is actually closer to f_3 than f_2.

A means of obtaining better estimates of peak density, etc., is outlined below. The technique is based on the fitting of a curve representing the response of a resonant circuit to the spectral estimates P_1, P_2, and P_3. The complex response of a maximum entropy filter near the region of small filter response (i.e., a spectral peak) is shown in Figure D-2. If Δf is much less than $1/T$, where T is the length of the filter, the complex response in the narrow frequency band near a point of minimum response can be approximated by a straight line. Then from solid geometry:

$$P_r(f) = d^2 + m^2 (f - f_o)^2.$$

Thus an approximation to the power spectrum near f_o is proportional to

$$P(f) \sim \frac{1}{P_r(f)} = \frac{1}{d^2 + m^2 (f - f_o)}$$

Defining:

$$a = 1/d^2 = \text{peak value}$$
$$b = 2d/m = 2/m\sqrt{a} = \text{bandwidth}$$

then we wish to fit the following curve to the spectrum

$$P(f) = \frac{a}{1 + 4/b^2 (f - f_o)^2} \qquad (D\text{-}1)$$

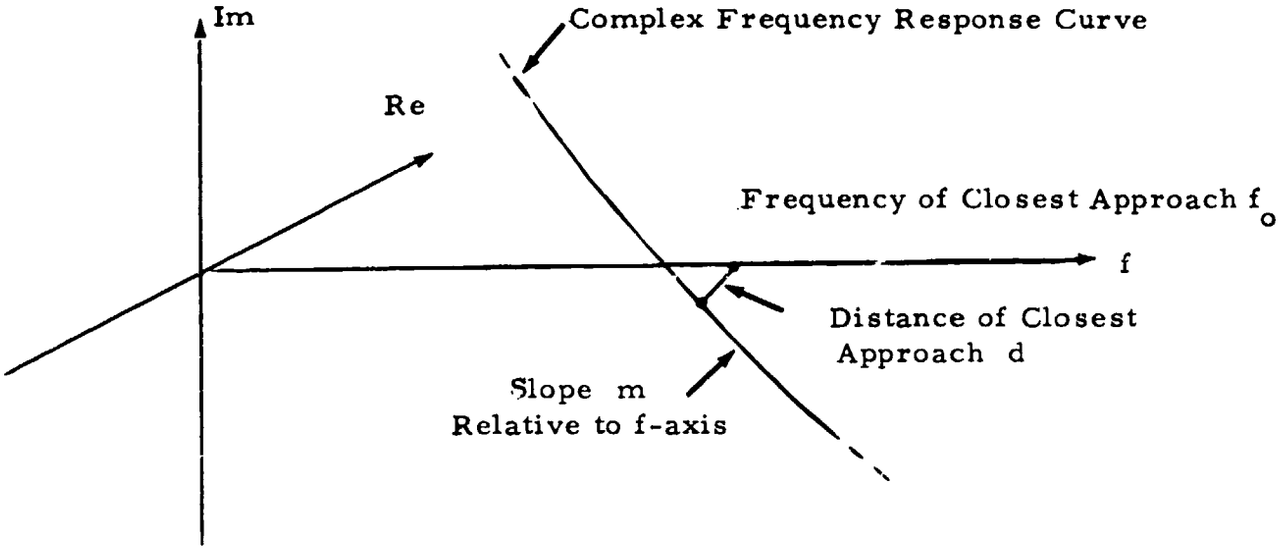


FIGURE D-2

COMPLEX FREQUENCY RESPONSE OF THE PREDICTION ERROR FILTER NEAR A MINIMUM

where:

$$P(f) = a \text{ for } f = f_o$$

$$= a/2 \text{ for } f = f_o \pm b/2$$

and the total power, T_p, is:

$$T_p = \int_{-\infty}^{\infty} P(f)\, df = \frac{\pi ab}{2} \quad . \tag{D-2}$$

Substituting the three values of power density about a peak ($P_1(f_1)$, $P_2(f_2)$, $P_3(f_3)$) into equation (D-1), the parameters a, b, and f_o can be solved for:

$$f_o = f_2 + Q\Delta f \tag{D-3}$$

$$a = P_2 \left[1 + \frac{Q^2}{\frac{2P_1}{R} - (1-Q)^2} \right] \tag{D-4}$$

$$b = 2\Delta f \left[\frac{2P_1}{R} - (1-Q)^2 \right]^{1/2} \tag{D-5}$$

where

$$R = P_1 + P_3 - \frac{2P_1P_3}{P_2}$$

$$Q = \frac{P_3 - P_1}{2R}$$

$$\Delta f = f_3 - f_2 = f_2 - f_1 \quad .$$

Occasionally, when the frequency interval Δf is too large, the quantity $(1 - Q)^2$ exceeds $2P_1/R$ in equation (D-5) causing imaginary solutions. A real solution can always be obtained, however, if Δf is sufficiently small. In that event, the spectrum is interpolated at the midpoints between P_1 and P_2 and P_3. This results in five power density estimates at $\Delta f/2$. The new peak density and its adjacent values are then used to obtain a solution. This interpolation procedure may be repeated as often as necessary. It is advantageous, however, to use a rather small Δf to start with so that the spectral density plots will give a fairly accurate picture of the true spectrum.

APPENDIX E

ADAPTIVE IMPLEMENTATION OF THE MAXIMUM ENTROPY SPECTRUM

The basic idea behind this technique is to smooth exponentially the prediction error filter output products used to estimate P_1 and the ladder coefficients a_{J+1}^{J+1} ($J = 1, 2, \ldots, N-1$) in the Burg technique (see Section III). Older prediction error filter output points are weighted slightly less than their immediate successors:

$$P_1 = \frac{\sum_{t=1}^{T} k^{T-t} x_t x_t^*}{\sum_{t=1}^{T} k^{T-t}} = \frac{U_1(T)}{V_1(T)} ;$$

$$a_{J+1}^{J+1} = \frac{-2 \sum_{t=J+1}^{T} k^{T-t} (p_{t-J}^J)^* q_t^J}{\sum_{t=J+1}^{T} k^{T-t} \left[p_{t-J}^J (p_{t-J}^J)^* + q_t^J (q_t^J)^* \right]} = \frac{U_{J+1}(T)}{V_{J+1}(T)} ,$$

where $0 < k < 1$.

The new prediction error filter outputs needed at time $t = T\Delta t$ are computed from the prediction error filter ladder coefficients obtained at $t = (T-1)\Delta t$. However, previously computed prediction error filter outputs used in forming the new prediction error filter outputs are not recomputed as the prediction error filter is updated. In this way, the number of arithmetic operations is made proportional to the filter length in points times the number

of time series input points. If the ladder coefficients change slowly, i.e., if $1-k << 1$, this procedure will have only a minor effect on the prediction error filter outputs and ladder coefficients.

The numerator and denominator terms $U_J(T)$ and $V_J(T)$ ($J = 1, 2, \ldots, N$) are recursively updated as each new time series value x_T becomes available, and the one-point prediction error filter output power and the ladder coefficients a_{J+1}^{J+1} ($J = 1, 2, \ldots, N-1$) are computed in the following manner:

$$P_1(T) = \frac{k\,U_1(T-1) + x_T x_T^*}{k\,V_1(T-1) + 1}$$

$$a_{J+1}^{J+1}(T) = \frac{k\,U_{J+1}(T-1) - 2(p_{T-J}^J)^*\, q_T^J}{k\,V_{J+1}(T-1) + \left[p_{T-J}^J (p_{T-J}^J)^* + q_T^J (q_T^J)^*\right]}.$$

Initially, the numerator and denominator terms $U_J(T)$ and $V_J(T)$ as well as the ladder coefficients $a_{J+1}^{J+1}(T)$ are zero at $T = 0$. During initialization ($T = 1, 2, \ldots, N$), only enough points of the time series x_t are available to compute the prediction error filter outputs for the T-point-long prediction error filter, so that the T-th ladder coefficient cannot be updated until x_T is available.

Figure E-1 is a flow chart of the procedure for recursively updating P_1 and the ladder coefficients a_{J+1}^{J+1}. Before the update at time $t = T\Delta t$, the processed time series is of the form $\ldots, p_{T-N-1}^N, p_{T-N}^N, p_{T-N+1}^{N-1}, p_{T-N+2}^{N-2}, \ldots, p_{T-3}^3, p_{T-2}^2, p_{T-1}^1 = x_{T-1}, x_T$. After the update, the processed time series is of the form $\ldots, p_{T-N-1}^N, p_{T-N}^N, p_{T-N+1}^N, p_{T-N+2}^{N-1}, \ldots, p_{T-3}^4, p_{T-2}^3, p_{T-1}^2, p_T^1 = x_T$. All of the backward prediction error filter output

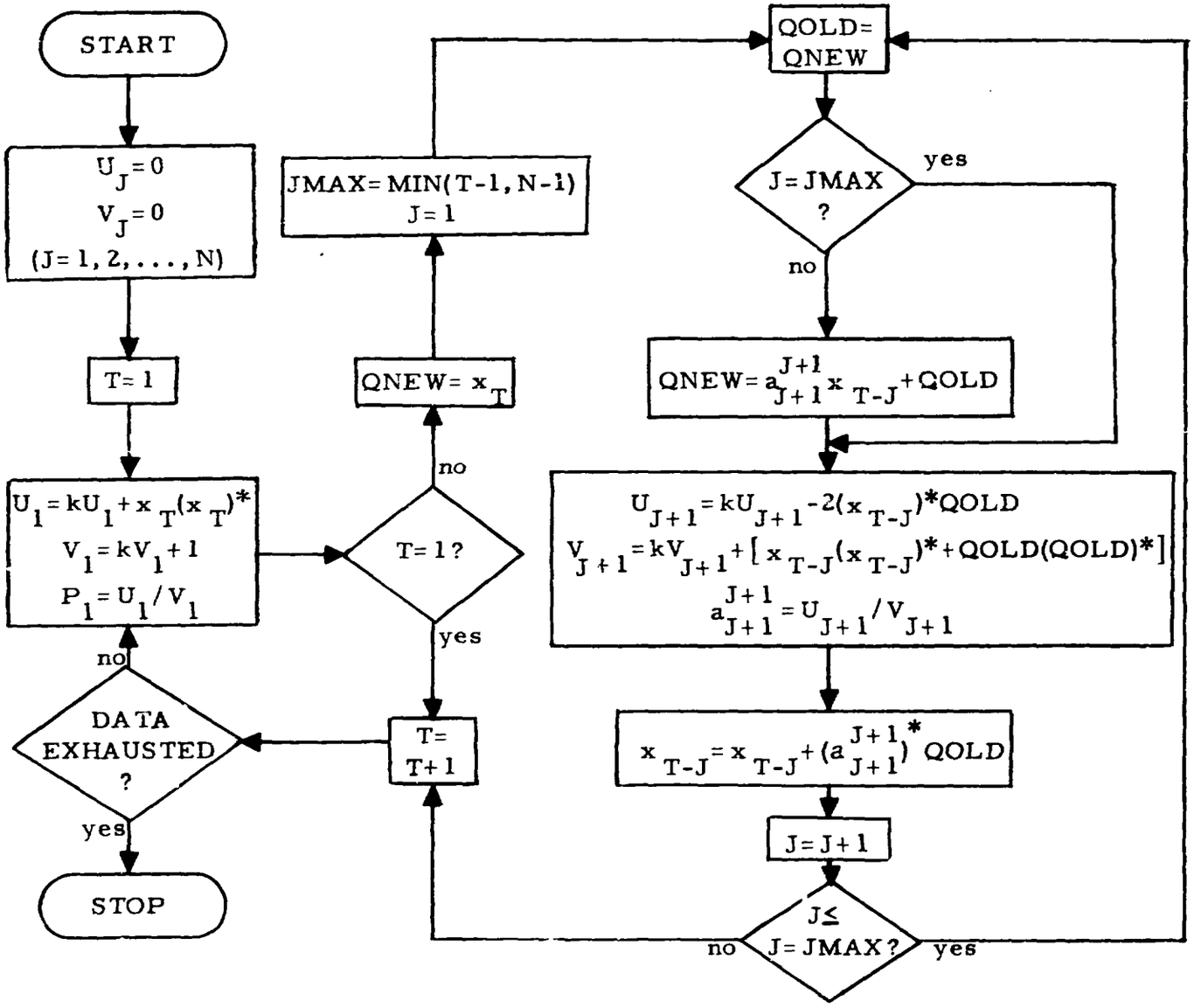


FIGURE E-1

PROCEDURE FOR RECURSIVE UPDATE OF DATA POWER AND PREDICTION ERROR FILTER LADDER COEFFICIENTS

points needed for the next update overlay the corresponding points of the input time series. The successive forward prediction filter outputs q_T^J ($J = 1, 2, \ldots, N-1$) at time $t = T\Delta t$ are computed as needed in the ladder coefficient update. Note that the backward prediction error filter outputs p_{T-J}^J and the forward prediction error filter outputs q_T^J in the ladder coefficient update are both computed from the prediction error filter obtained during the update at time $t = (T-1)\Delta t$. The recursive procedure described here has been outlined previously in general terms (Riley and Burg, 1972) but not explicitly stated and differs from the technique actually implemented by Burg (Burg, 1974).

To compute a maximum entropy spectrum, the Levinson recursion relations

$$\begin{bmatrix} 1 \\ \hline a_2^{J+1} \\ \vdots \\ a_J^{J+1} \\ \hline a_{J+1}^{J+1} \end{bmatrix} = \begin{bmatrix} 1 \\ \hline a_2^{J} \\ \vdots \\ a_J^{J} \\ \hline 0 \end{bmatrix} + a_{J+1}^{J+1} \begin{bmatrix} 0 \\ \hline (a_J^J)^* \\ \vdots \\ (a_2^J)^* \\ \hline 1 \end{bmatrix} \qquad (J = 1, 2, \ldots, N-1)$$

and

$$P_N = P_1 \prod_{J=1}^{N-1} \left[1 - a_{J+1}^{J+1} \, (a_{J+1}^{J+1})^* \right]$$

provide the N-point-long prediction error power P_N and the prediction error filter $(1, a_2^N, a_3^N, \ldots, a_{N-1}^N, a_N^N)$ needed in the power density spectrum formula

$$P(f) = \frac{P_N \Delta t}{\left| 1 + \sum_{J=1}^{N-1} a_{J+1}^N \, e^{i2\pi f J \Delta t} \right|^2} .$$

The power density spectrum requires considerably more computational effort than a single update of P_1 and the ladder coefficients. If desired, the power density spectrum can be computed only at specified time intervals instead of after every update in order to reduce the computational load.

If $k = \tau/(\tau+1)$, $k^{-m} = e$ when $m = 1/\ln[(\tau+1)/\tau]$ or approximately $\tau + 0.5$ when τ is large, so that the time constant for the adaptive maximum entropy spectrum is approximately $(\tau + 0.5)\Delta t$. After M updates, the number of degrees of freedom in estimating the power P_1 is

$$\sum_{m=1}^{M} k^{M-1} = \frac{1 - k^M}{1 - k} .$$

As M becomes large, this value approaches $1/(1-k)$ or $\tau + 1$ if $k = \tau/(\tau+1)$. The effective time delay is

$$\frac{-\Delta t \sum_{m=1}^{M} (m-1)k^{M-1}}{\sum_{m=1}^{M} k^{M-1}} = \frac{-k\Delta t}{1 - k^M}\left[\frac{1 - k^{M-1}}{1 - k^M} - (M-1)k^{M-1}\right]$$

after M updates and approaches $-k\Delta t/(1-k)$ or $-\tau\Delta t$ as M becomes large.

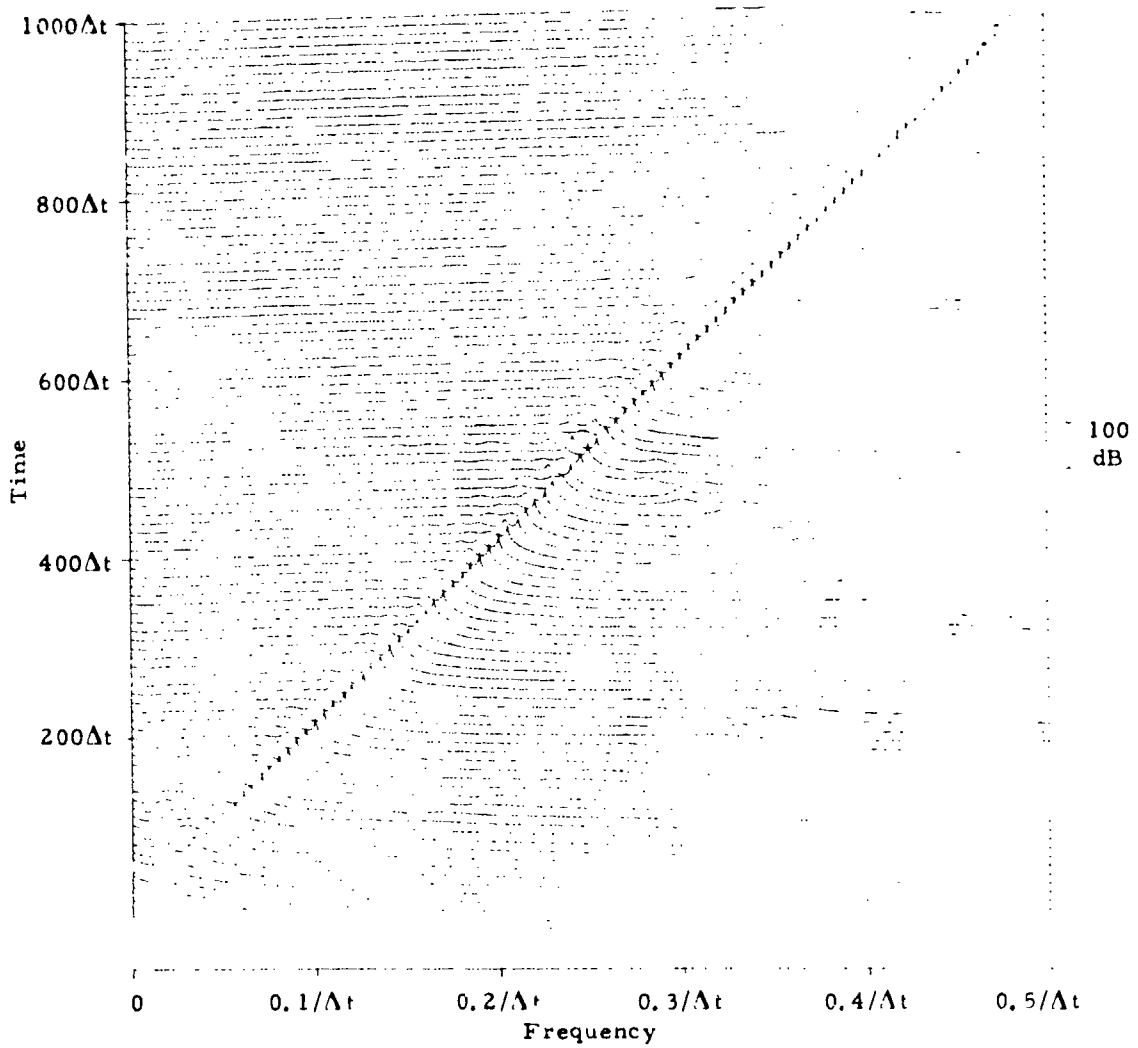
To test the adaptive maximum entropy spectral analysis technique described here, a 1000-point-long chirp waveform

$$x_T = x(T\Delta t) = \sin\left[\frac{\pi (T-1)^2}{2TMAX}\right] \qquad (T = 1, 2, \ldots, TMAX)$$

was processed. The instantaneous frequency of this waveform is

$$f(T\Delta t) = \frac{T-1}{2TMAX\Delta t} \quad .$$

The length of the prediction error filter was fifty points. The time constant used in the recursive adaptive update was $100\Delta t$, or twice the prediction error filter length. After every ten points, the maximum entropy spectrum corresponding to the current prediction error filter and the current prediction error power was computed over the frequency band 0 to $0.5/\Delta t$ at a frequency increment of $0.0005/\Delta t$. Figure E-2 displays the resulting adaptive power density spectra in logarithmic form. As time increases, corresponding spectral levels are plotted at a higher level in the figure. After a brief warmup period, the principal spectral peak closely follows the instantaneous frequency of the input time series. Because of the exponential smoothing applied to the input data, the trend of the spectrum is a linear increase (on a logarithmic scale) as the frequency rises to the frequency of the principal spectral peak and then a sharp dropoff out to the folding frequency. Figure E-3 illustrates the ability of the adaptive maximum entropy spectrum to track the instantaneous frequency of the chirp waveform. In this figure, the frequency of the maximum spectral intensity is plotted as a function of time. Except for a small number of points, the resulting path is very close to linear. The time lag of the spectral peak relative to the instantaneous frequency is about ten points, or only one tenth of the time constant.



$$x(T\Delta t) = \sin\left[\frac{\pi(T-1)^2}{2TMAX}\right] \qquad (T = 1, 2, \ldots, TMAX)$$

$$f(T\Delta t) = \frac{T-1}{2TMAX\Delta t}$$

Length of Prediction Error Filter = 50 points

Time Constant = 100Δt

FIGURE E-2

ADAPTIVE MAXIMUM ENTROPY SPECTRUM OF CHIRP WAVEFORM

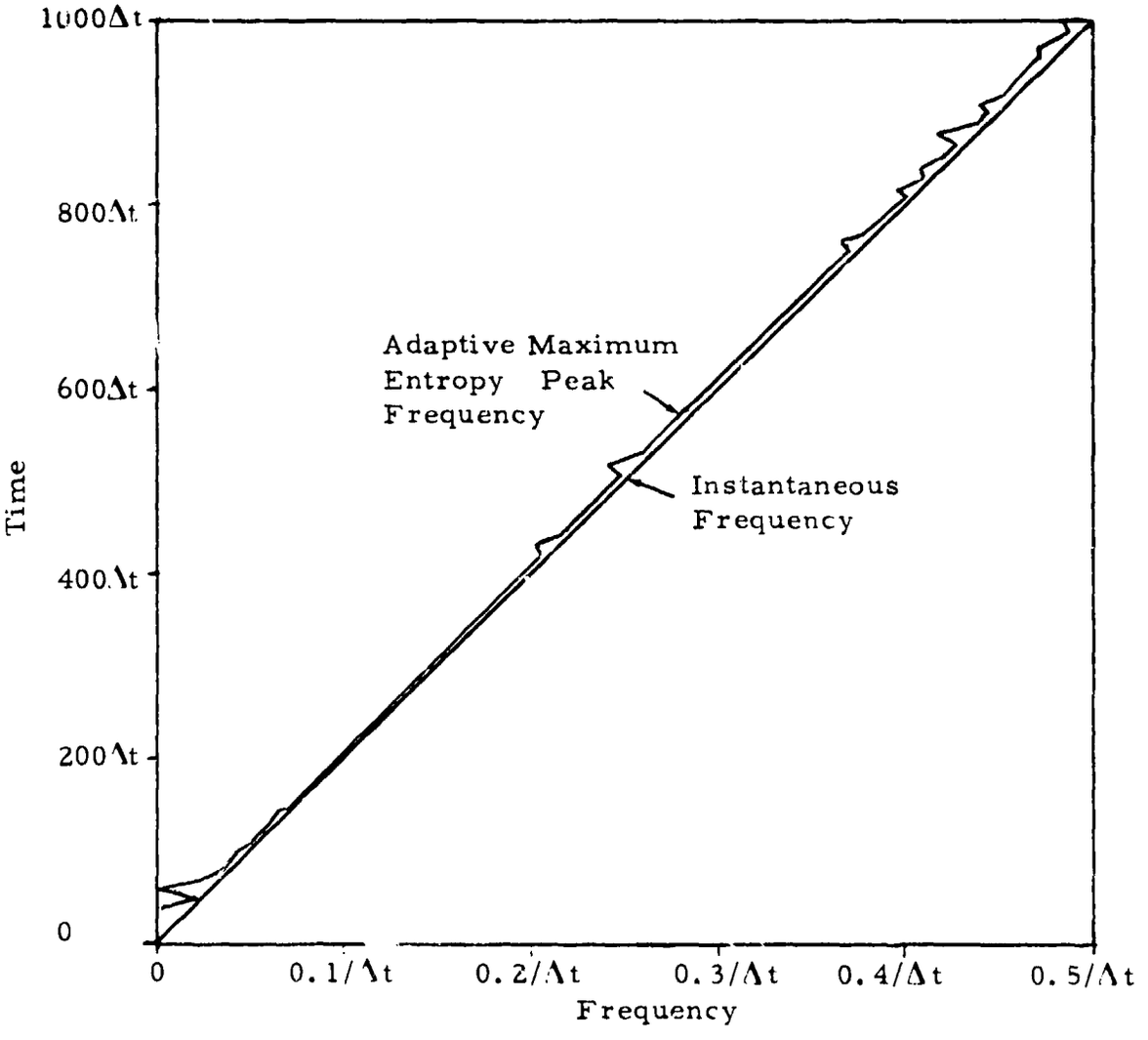


FIGURE E-3

FREQUENCY OF MAXIMUM SPECTRAL INTENSITY AS A FUNCTION OF TIME FOR ADAPTIVE MAXIMUM ENTROPY SPECTRUM
(Time Constant = 100Δt)